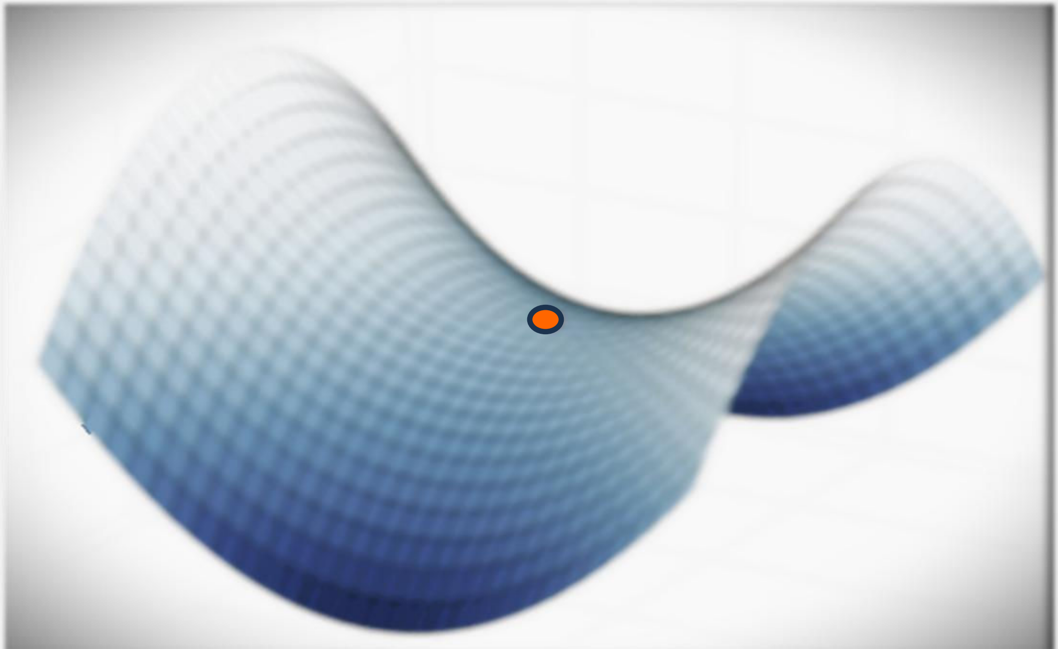
AI Risk Assessment: A Central Governance Pillar

ValidMind & Experian
AI Risk Summit`25, December 2, 2025

Rodanthy Tzani, Ph.D.
Founder, Risk & Compliance Advisor

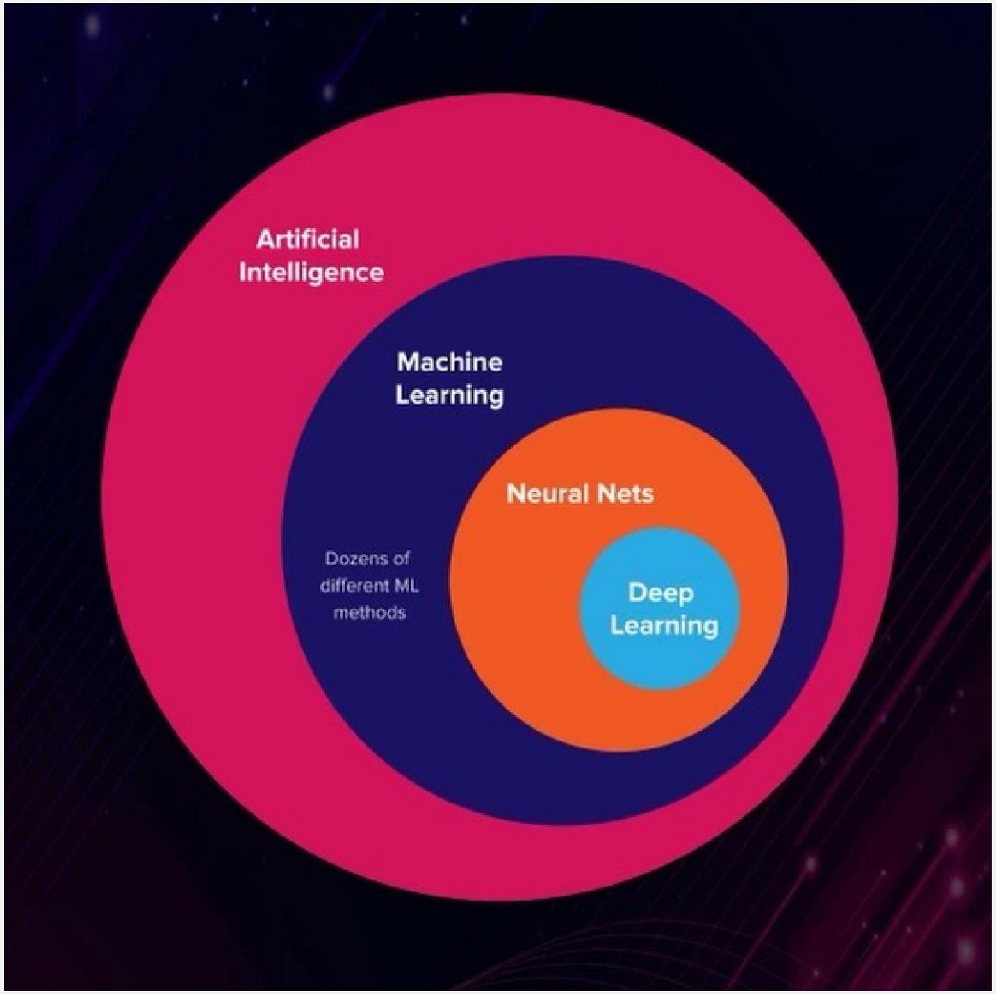
Sphaleron



Agenda

1. AI Applications Definition
2. Risks of AI & GenAI Applications
3. Regulatory Landscape in Insurance
4. Overview of AI Risk Management Framework
5. AI Inventory
6. AI Lifecycle & Risk Assessments
7. Conclusion

Artificial Intelligence (AI) Applications Definition



- **AI Applications** are models, systems or programs that utilize AI technologies, including Machine Learning, Neural Network Processing, Deep Learning, and train on data to explain or predict outcomes

- ML is a subset of AI: process of **training algorithms** on existing data to learn patterns and make predictions and decisions

- Generative AI encompasses a **broader range of technologies**, **algorithm**s or **advanced techniques** that enable machines to perform tasks that would typically **require human intelligence**

Traditional AI Risks Proliferated and Accelerated by GenAI

Key Risks of AI/ML Models

- **Faster model development** cycles due to availability of advanced analytical tools/platforms introduce the potential risk of deploying **complex models** without fully **understanding underlying risks**

- While better predictions **improve efficiency** and effectiveness of key processes, lack of **interpretability or explainability** makes it difficult to attribute output to key factors

- Potential **to improve customer experience** & expand financial inclusion increases concerns about **unintended bias** due to alternative data and complexity of algorithms

- **Regulatory, reputational** and **potential financial risk** due to possible violation of **antidiscrimination** laws

Additional Risks of GenAI Use Cases

- **Lack of transparency**, **speed**, and **volume of decisions** made by AI pose challenges; need to ensure appropriate level **of human accountability**

- AI systems are built to be efficient and effective for a defined purpose, but using them may not take-into-account **ethical values, legal context**, or other trade-offs

- AI systems are designed to **improve speed of decision making**. This can result in a rapid and wide scale **harm to external stakeholders** and/or the company, ultimately leading to **legal, reputational** and/or **financial losses**

- **Big and untested data** are used as inputs; may result to issues with **data privacy, reliability, integrity, and relevance.**

U.S. Regulatory Landscape for AI in Insurance

❖Insurance is state-based regulation - in progress and evolving

As of July 2025

← **Regulations** —— **High-Level Expectations** →

Principle-based

- ➢ Principles - adopted in 2020
- ➢ **Model Bulletin**[1] for Insurers adopted Dec. 4, 2023

Key AI principles:

1. Fairness and equity; 2. Accountability
3. Transparency; 4. Compliance with laws
5. Security, Safety, Robustness

Colorado DOI

Prescriptive

- ➢ **Regulation 10-1-1 -** Effect. Nov. 14, '23
- ➢ Amendment extends regulation to **private passenger auto** and **health benefit plan** insurers – Effective June 5, 2025

Specific Requirements:

Documented **Governing Principles** & Policies
Board and Senior Mgt oversight & accountability Supervision and training for employees
Inventory of AI models
Bias testing to detect unfair **discrimination**

Industry Guidance

- ➢ NYS DFS[3] published the **Circular Letter** Insurers that utilize (ECDIS) for underwriting & pricing models – published July 2024

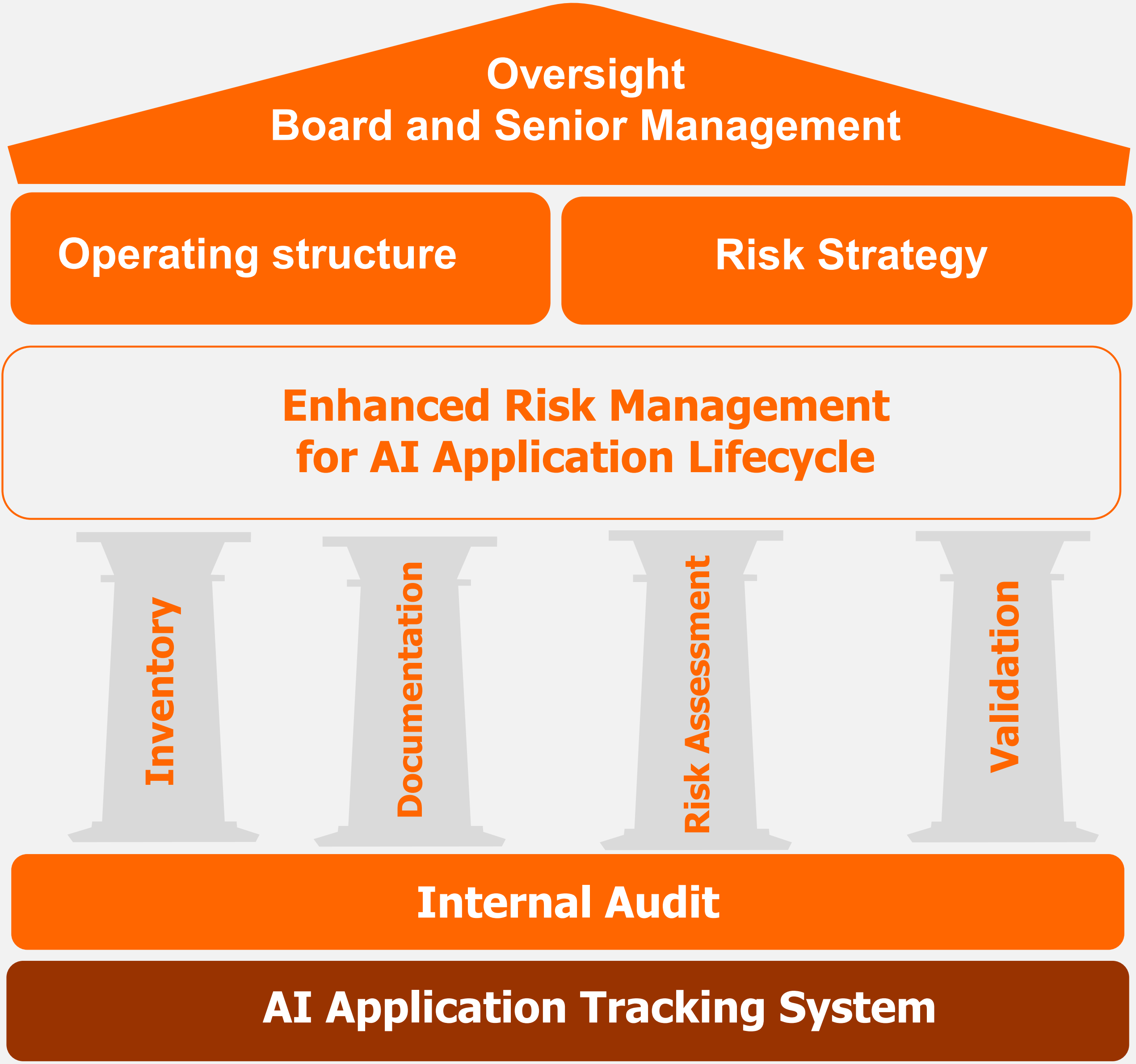
Expectations:

Formalize **Governance & Risk Management** of AIS (& ECDIS) in policies and procedures
Board & Senior Mgt oversight of AIS & ECDIs governance, including Third-Party
Test data for **discriminatory bias**

[1] Based on the "Unfair Trade Practices Model Act (#880)" and the "Unfair Claims Settlement Practices Model Act (#900)"
[2] External Consumer Data and Information Sources
[3] New York State Department of Financial Services

❑ AI Applications Require Enhanced Risk Management Framework and Effective Controls

AI Application Risk Management Framework (I/II)



Oversight
Board and Senior Management

Operating structure

Risk Strategy

Enhanced Risk Management for AI Application Lifecycle

Inventory

Documentation

Risk Assessment

Validation

Internal Audit

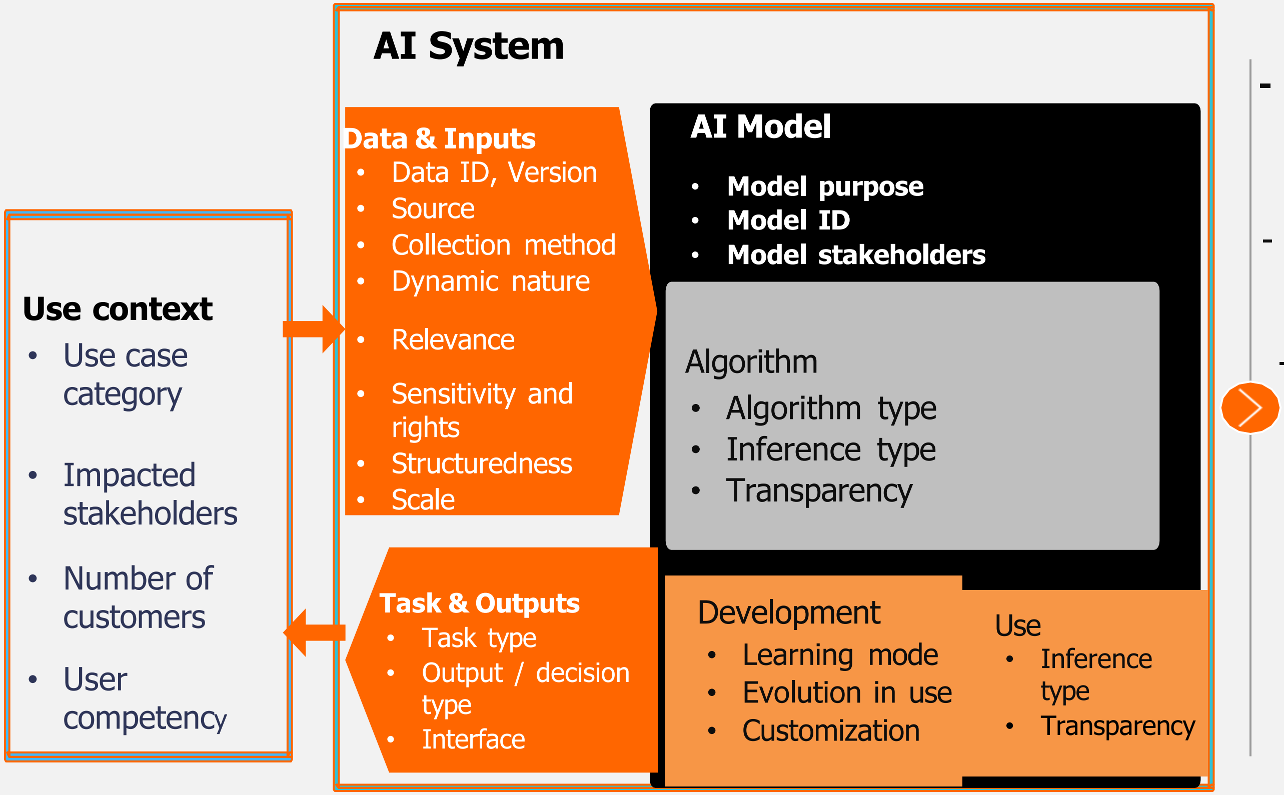
AI Application Tracking System

AI Application Risk Management Framework (II/II)

Components	Summary
Oversight Enhanced	• **Management Oversight:** Board, Risk Steering Committee • **Defined Roles and Responsibilities** within the Three Lines of Defense model • **Established Metrics** of AI risk and related risk events
Components of Governance Enhanced	• **Operating Structure: Cross functional governance** committee; AI policy & standards; Enhanced roles & responsibilities; controls for AI lifecycle; • **Documentation: AI** development standards, AI implementation, use and validation • **AI Inventory:** AI definition and identification across company; AI in tracking system; version control • **AI Risk Assessment:** identification and quantification of risk factors; method for risk assessment; risk tiering; individual and aggregate risk metrics • **Plan of Validation and Prioritization:** process for prioritizing for validation • **Internal Controls & Process for AI Approval:** using AI that are not reviewed or validated
Reporting & Tracking System Enhanced	• **Reporting and Communication: C**ommunication across all three lines of defense • **Risk Technology:** Enterprise tracking system and risk governance tool • **Education and Training:** implement training for all stakeholders
Validation & Monitoring Enhanced	• **Validation Practice for AI Lifecycle:** risk-based independent validation; effective challenge; on-going monitoring

AI Inventory

AI Application Inventory Captures Key Risk Attributes of the System as a Whole



Use context

- Use case category
- Impacted stakeholders
- Number of customers
- User competency

AI System

Data & Inputs

- Data ID, Version
- Source
- Collection method
- Dynamic nature
- Relevance
- Sensitivity and rights
- Structuredness
- Scale

Task & Outputs

- Task type
- Output / decision type
- Interface

AI Model

- **Model purpose**
- **Model ID**
- **Model stakeholders**

Algorithm

- Algorithm type
- Inference type
- Transparency

Development

- Learning mode
- Evolution in use
- Customization

Use

- Inference type
- Transparency

AI Inventory facilitates:

- Governance and Oversight
- Accountability
- Risk management
- Lifecycle management
- Reporting
- Regulatory compliance

Source: Adapted from: BCG/Microsoft Responsible AI Guidelines, OECD AI systems classification; ALTAI self-assessment; EIOPA AI Governance; NAIC Life insurance survey

AI Inventory Includes Attributes Ensuring Comprehensiveness and Usability

Illustrative

Enhance existing model inventory attributes

Basic identification

- Basic information for tracking and identification of AI applications; ID, name, description
- IDs of owner, developer, user and uses
- Developing platform, assumptions, inputs, data used for development
- Implementation platform and techniques, production data
- Model/Algo dependences
- If third party/vendor model; vendor information
- Materiality, Exposure metrics
- Usage and model type

Governance aspects

- Approval status
- Attestation status
- Attestation owner
- Version ID
- Risk ratings for inherent and residual risk
- Validation status; findings

Incorporate AI related characteristics

- Is this an AI model or GenAI use case?
- Use context; Is there potential for harm?
- Is there a human in the loop?
- Is this a customer facing solution?
- Impacted stakeholders
- Methodology; AI techniques
- Development platform
- Size, exposure metrics
- Additional AI Vendor/Third Party information, if applicable
- Data information
- Algorithms and predictive models that utilize ECDIS[1]
- Regulatory expectations, if any
- Enhanced risk assessments for inherent and residual risks
- Bias metrics

Enhanced

[1] ECDIS: External Consumer Data and Information Sources

AI Lifecycle and **Risk Assessments**

Define Level of AI Risk

- Use Existing Regulation, Company's Internal Principles & External Ethical, Fairness, Environmental and Societal Concerns

Non-exhaustive

Prohibited/Unacceptable

- Cognitive behavioral manipulation of people
- Social scoring AI; e.g., classifying people based on socio-economic status
- Biometric categorization of people
- Automation of potential harmful decisions
- Use for surveillance or tracking
- Use of personal data without consent

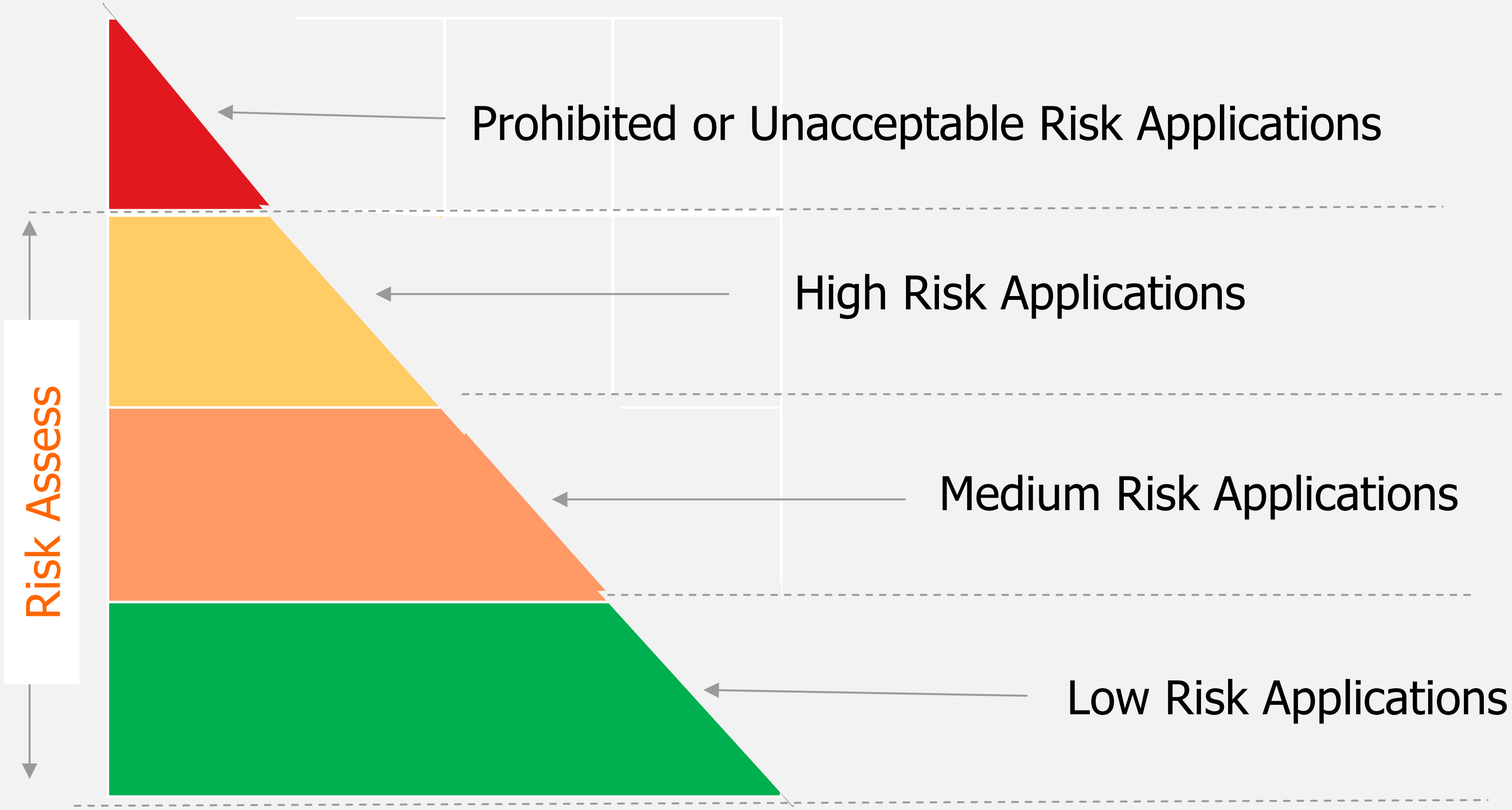
High Vigilance

- Management and operation of critical infrastructure
- Core business for decision making
- Employment decisions, worker management
- Data, external consumer data (ECDIS)*
- Law enforcement
- Education and vocational training

* ECDIS = External Customer Data Information Sources

Initial Risk Framework to Address Regulatory and Other Principles

❖ To be monitored and evaluated over time



Prohibited or Unacceptable Risk Applications

High Risk Applications

Medium Risk Applications

Low Risk Applications

Inherent and Residual AI Risk Assessments to be Implemented along the AI Lifecycle

Ideation

AI Risk Screening

- Risk Screening to
- identify **prohibited vs high-vigilance** AI use cases; review against **Risk Appetite**

- Identify major potential **harms.**
- Consider potential **legal and regulatory** constraints

- Classify use cases according to **inherent risk**: Low, Medium, High

Design

AI Risk Assessing

- **Deep risk assessment** for all AI use cases flagged for during the risk screening

- Produce **inherent risk scoring** for use cases and recommend **risk mitigating** actions

- Ensure alignment with **AI principles**

Develop & Implement

AI Risk Evaluation and Validation

- Risks associated with **development** process - Introduce **mitigating** controls

- Perform **risk-based independent validation** & assess the effectiveness of risk mitigation measures

- **Complete risk assessment.** Initiate **risk acceptance** process based on **residual risk level**

Use

AI Risk Monitoring

- **Risk assessment & performance evaluation** during ongoing usage

- **Monitor risk** profile of use case, identify new risks, and monitor mitigation process

- Monitor changes in **legal** and **regulatory requirements**.
- Identify anomalous behaviors

1

Identify risk attributes of AI applications are aligned with key risk dimensions:

- Severity of Impact
- Likelihood of Failure

AI Inherent Risk Approach

Materiality:
Measures financial significance of the business area and the process where the AI will be used

- Impact on financials
- Impact on individuals

Importance/Criticality:
Measures severity of the negative impact if AI fails

- Business context, etc.
- Individuals' potential harm/discrimination, etc.

1 **Severity of Impact**

Context of the AI application that results in a more severe impact upon its failure

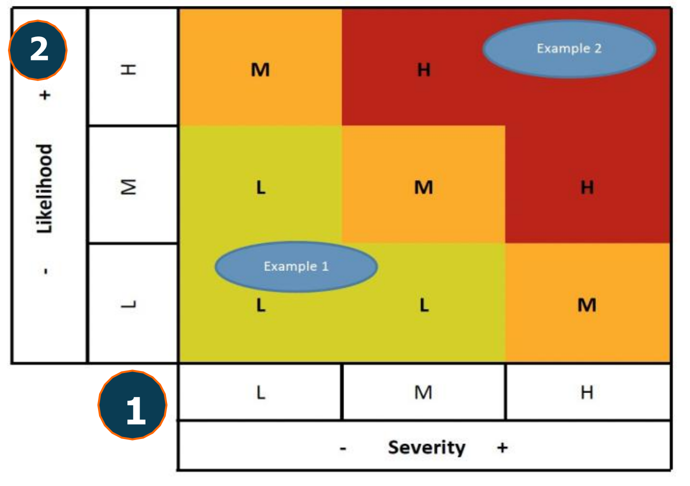
2 &

Likelihood of Failure

Aspects of the AI application that make it more likely to fail

- Measures how likely a failure to occur given the characteristics of the AI application

Assessment Matrix



2 Likelihood				
+	H	M	H	Example 2
	M	L	M	H
-	L	Example 1 L	L	M
1		L	M	H
		- Severity +		

🟥	High Risk
🟨	Medium Risk
🟩	Low Risk

2

Collect data to proxy the risk attributes identified in 1. Adjust the data for accuracy and consistency

Key Questions to Identify Risk Factors and Collect Data

Examples

- □ **Who is impacted** by the AI application or business solution, either directly or indirectly?
- □ **How might** the AI application **fail to perform** as intended?
- □ What are the **potential harms** to impacted individuals if the AI application fails?
- □ Are there any additional **legal, regulatory**, or **policy implications** associated with the identified harms?
- □ Is this AI application **automated & making** potentially harmful decisions?
- □ Is this AI application using **personal data** without consent?

Questions to be answered by SMEs

3 Defining Risk Factors and Combining them to Assess Dimensions of Risk

Illustrative

1 Severity of Impact

Materiality factors:

- **Impact on financials:**
  - □ Financial/Business exposure
- **Impact on individuals:**
  - □ Number of stakeholders

Importance/Criticality factors:

- **Business context:**
  - □ Business area (Business context purpose)
  - □ Task type (Optimization, Classification, Event detection, Forecasting, Personalization)
  - □ Legal and regulatory requirements
  - □ Reputational impact
  - □ Business continuity
- **Individuals’ potential harm:**
  - □ Type of stakeholders (Employees, Customers)
  - □ Potential for harm (Adverse decisions
  - □ Potential for harming Individuals/Groups

2 Likelihood of Failure

Risk of Failure factors:

- **Data:**
  - □ Relevance to decision
  - □ Sensitivity level (employee or customer ID)
  - □ Source (Internal vs. Vendor vs. Open)
  - □ Structured vs non-Structured

Model:

- □ Internal or Third-Party
- □ Model and dataset complexity (Number of Model Components, Datasets)
- □ Algorithm type (Discriminative, Generative)
- □ Training model (Supervised, Unsupervised)
- □ Model dependency (Upstream and Downstream)

- **Output/ Decision:**
  - □ Inference type (Deterministic vs. Probabilistic)
  - □ Level of decision autonomy (Automation vs. Augmentation)
- **Implementation:**
  - □ Internal vs. Vendor
- **Interface:**
  - □ Internally vs. Externally facing

Biggest drivers result to financial losses & reputational impacts

4 Calibrate the Risk Assessment method based on high-vigilance areas and AI Risk Appetite

Illustrative

Define the weight distribution among risk dimensions

1. Equal weights

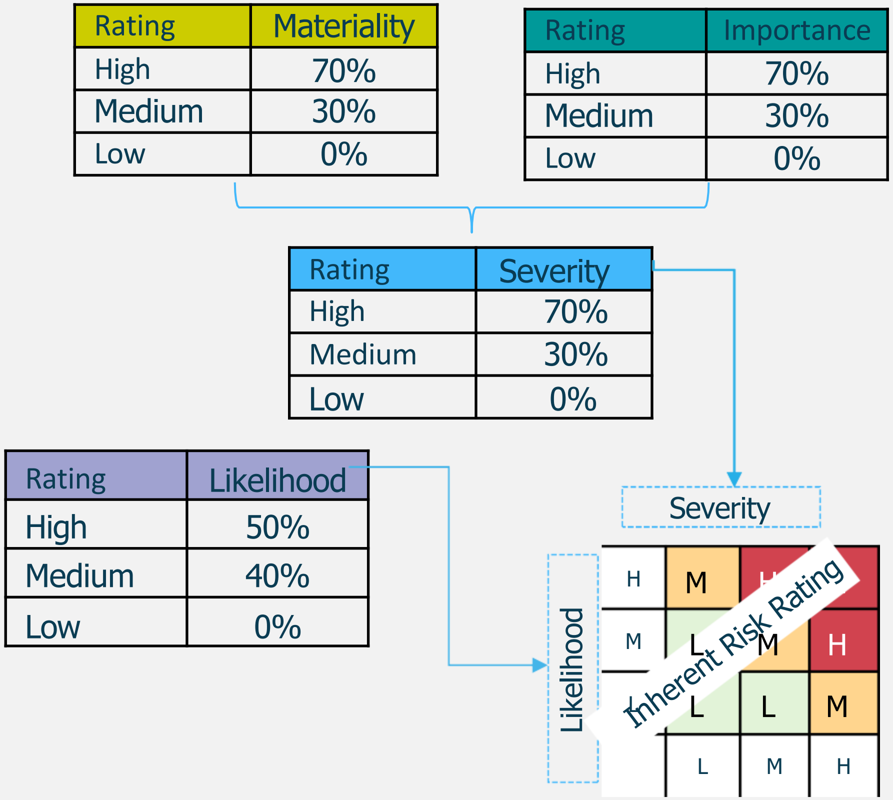
2. Overweight some components

- Higher weights for some factors (e.g., data sensitivity, automated decision-making and customer interface)
- Higher importance in alignment with regulatory guidelines etc.

3. Conditional approach

- Use max values for certain dimensions (e.g., data, automation)

Define thresholds for level of risks for each risk dimension



Rating	Materiality
High	70%
Medium	30%
Low	0%

Rating	Importance
High	70%
Medium	30%
Low	0%

Rating	Severity
High	70%
Medium	30%
Low	0%

Rating	Likelihood
High	50%
Medium	40%
Low	0%

Severity

Likelihood

H	M	H	[illegible]
M	L	M	H
L	L	L	M
	L	M	H

Inherent Risk Rating

Explore the impact of alternative weights and thresholds to identify high-risk areas

- Q & A

Sphaleron