

---

# The Enhanced 4-Lenses Framework for AI Risk Classification

A Structured Instrument for Understanding AI Risk

---

Jan W. Veldsink MSc

# The Classification Challenge

Traditional frameworks classify IT systems. AI systems require a different approach.

The **AIC triad** (Availability, Integrity, Confidentiality) was designed for predictable infrastructure. AI systems present fundamentally different risk characteristics that traditional classification cannot adequately capture.

AI systems are fundamentally different:



## Autonomous Decision-Making

Systems act independently without direct human intervention



## Learning and Adaptation

Behavior changes over time in ways that are difficult to predict



## Opacity and Complexity

Decision processes create "black box" challenges in understanding

# From AIC to 4-Lenses

The evolution from traditional to AI-specific risk classification

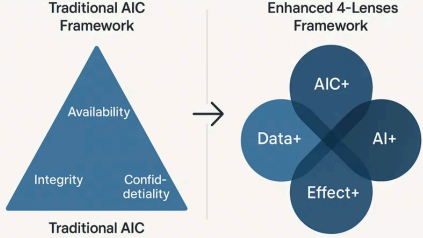
**Executive Summary: Enhanced 4-Lenses Framework for AI Governance**

Author: Jan W Veldsink MSc  
Date: July 17 2025

**The Challenge**

Traditional information security frameworks, particularly the Availability, Integrity, and Confidentiality (AIC) triad that organizations have relied upon for decades, are fundamentally inadequate for governing artificial intelligence systems. The emergence of autonomous AI agents, foundation models, and adaptive learning systems has created governance gaps that expose organizations to unprecedented business, regulatory, and reputational risks.

**Why AIC Alone Is Insufficient for AI Governance**



AIC vs Enhanced 4-Lenses Comparison

**Why Traditional AIC Frameworks Are Insufficient for AI Governance**

The European Union's AI Act, which entered force in August 2024, represents the first comprehensive regulatory framework for AI systems globally, establishing risk-based classifications that extend far beyond traditional security considerations. Similar regulatory developments across multiple jurisdictions indicate a global shift toward AI-specific governance requirements that conventional IT security frameworks cannot adequately address.

1

AI risks require classification through four integrated lenses beyond AIC.

# What Is a Classification Instrument?

The framework helps you see and understand AI risk, not prescribe what to do about it

## Classification

Systematic categorization of AI systems based on shared risk characteristics across four dimensions

## Diagnostic, Not Prescriptive

The framework **identifies and classifies** risks but does not dictate specific controls or governance mechanisms

## Risk Assessment

Evaluation of potential consequences through structured questions that reveal the AI risk profile

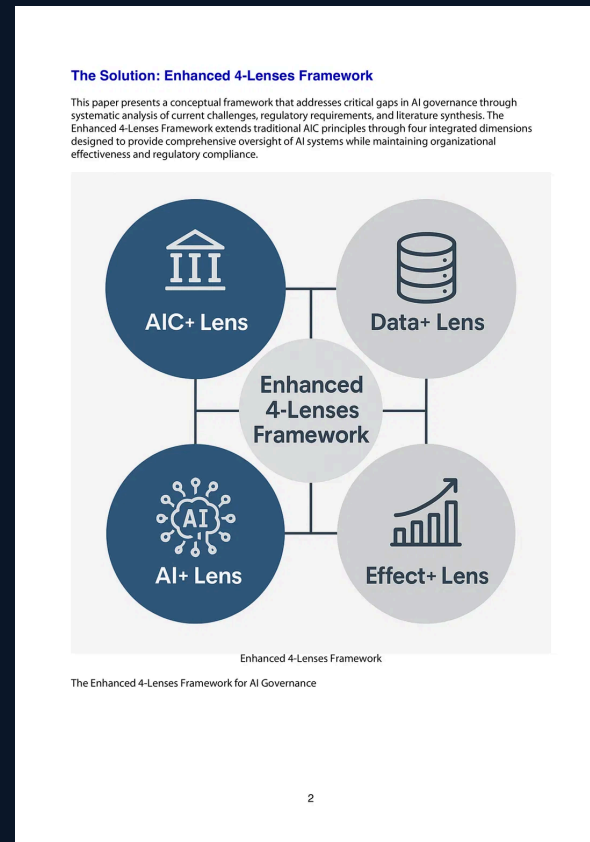
## Output: Understanding

Comprehensive risk profile that enables **informed governance decisions** based on actual risk characteristics

Like Cynefin helps you understand complexity, the 4-Lenses Framework helps you understand AI risk

# The Four Lenses Overview

Four integrated perspectives for comprehensive AI risk classification



**1** **AIC+**  
Enhanced Security Classification

**3** **AI+**  
AI System Risk Classification

**2** **Data+**  
Dynamic Data Risk Classification

**4** **Effect+**  
Impact Risk Classification

# Lens 1: AIC+ Enhanced Security Classification

**Purpose:** Classify AI-specific security risks that extend beyond traditional availability, integrity, and confidentiality concerns. Output: Security risk profile accounting for AI's autonomy, learning, and opacity.

## How Each Dimension Extends to AI

### Availability

#### Traditional AIC

Is the system available?

#### AIC+ for AI

How does autonomous agent behavior affect availability? Agent resilience, not just server uptime.

### Integrity

#### Traditional AIC

Is the data intact and unmodified?

#### AIC+ for AI

What model integrity risks exist? Adversarial attacks, model tampering, training data poisoning.

### Confidentiality

#### Traditional AIC

Is access to data controlled?

#### AIC+ for AI

What confidentiality challenges arise from agent communication or federated learning?

# AIC+ in Practice

## SCENARIO

Multi-agent customer service system with autonomous agents handling inquiries, escalations, and knowledge sharing

## Traditional AIC Protects

### Server Infrastructure

System uptime and availability monitoring

### Database Integrity

Data consistency and backup procedures

### Access Control

User authentication and authorization

## AIC+ Also Classifies

### Agent Communication Security

Risks in agent-to-agent information exchange and coordination

### Model Tampering Detection

Adversarial attacks, model version control, training data poisoning

### Privacy-Preserving Learning

Confidentiality in learning from customer interactions

### Autonomous Behavior Boundaries

Agent decision authority limits and escalation protocols

# Lens 2: Data+ Dynamic Data Risk Classification

**Purpose:** Classify data risks that change dynamically based on how AI systems use that data. Output: Dynamic data risk profile reflecting multiple usage contexts throughout the AI lifecycle.

## The Fundamental Difference

### Traditional Data Classification

Data is classified once and that classification remains static

### Data+ for AI

Risk profile changes as data moves through training, inference, and feedback contexts

## Risk Changes Across AI Usage Contexts

### Training Data

Bias detection, provenance tracking, quality assurance

### Vector Embeddings

New forms of sensitive information in unstructured representations

### Feedback Loops

Data quality degradation, drift detection over time

# Data+ in Practice

## SCENARIO

Foundation model training using diverse data sources for customer service automation

## Traditional Data Governance

### Source Data Classification

Static classification of original data sources

### Access Controls

Role-based permissions for data access

### Compliance Checks

Regulatory compliance at data collection

## Data+ Also Classifies

### Bias Detection Across Datasets

Systematic bias assessment across diverse training data sources

### Provenance Tracking

Data lineage for model behavior explanation and accountability

### Quality Assurance Throughout Lifecycle

Data quality monitoring from training through inference and feedback

### Context-Aware Classification

Risk profile changes as data moves between training, fine-tuning, and inference contexts

# Lens 3: AI+ AI System Risk Classification

**Purpose:** Classify risks inherent to AI models, agents, and systems themselves. Output: AI-specific risk classification based on system architecture, autonomy, and capability characteristics.

## Key Classification Questions

### Autonomy Level

What is the autonomy level of this system, and what risks does that create?

- Recommendation systems
- Assisted decision-making
- Autonomous agents

### Capability Drift

What capability drift or performance degradation risks exist over time?

- Performance degradation
- Unexpected capability emergence
- Model behavior changes

### Foundation Model Risks

What foundation model risks apply to this system?

- Capability assessment
- Safety boundaries
- Emergent behaviors

# AI+ in Practice

## SCENARIO

Autonomous decision-making agent for supply chain optimization with authority to adjust inventory levels and supplier orders

## Traditional Risk Assessment

### Application Layer Security

Standard software vulnerability assessment

### Business Logic Validation

Rule-based decision verification

### System Integration

API security and data flow controls

## AI+ Also Classifies

### Agent Autonomy Level

Decision authority boundaries and escalation thresholds for autonomous actions

### Capability Assessment

Foundation model capabilities and safety boundaries for decision-making

### Performance Drift Detection

Monitoring for capability degradation or unexpected behavior emergence

### Multi-Agent Coordination

Emergent behaviors from agent interactions and distributed decision-making

# Lens 4: Effect+ Impact Risk Classification

**Purpose:** Classify real-world impact risks that AI systems create for organizations, stakeholders, and society. Output: Impact risk profile accounting for intended and unintended consequences across stakeholder groups.

## Key Classification Questions

### Consequences

What are the potential consequences of autonomous decisions made by this system?

### Measurement

How do we measure and validate the actual impact of AI-driven actions?

### Stakeholder Risks

What organizational, cultural, and stakeholder risks does this AI usage present?

## Impact Risk Examples

### Hiring Decisions

Bias impact on candidate populations and organizational culture

### Predictive Maintenance

Operational impact of false positives and negatives

### Customer Service Agents

Brand impact of autonomous interactions and escalations

# Effect+ in Practice

## SCENARIO

AI-powered hiring system that screens resumes, ranks candidates, and recommends interview selections

## Traditional Impact Assessment

### Process Efficiency

Time savings in resume screening and candidate evaluation

### Cost Reduction

Reduced HR workload and recruitment costs

### Compliance Verification

Legal requirements for non-discrimination

## Effect+ Also Classifies

### Bias Impact on Candidates

Systematic exclusion of qualified candidates from underrepresented groups

### Organizational Culture Effects

Long-term impact on team diversity, innovation, and workplace dynamics

### Stakeholder Trust and Brand

Reputation risk from perceived unfairness or discriminatory practices

### Outcome Measurement

Validation of actual hiring quality and performance versus AI predictions

# Integration: The Four Lenses Work Together

Comprehensive risk profiles emerge from examining AI systems through all four integrated lenses

## AIC+

Reveals AI-specific security risks in autonomy, model integrity, and agent communication

## Data+

Exposes dynamic data risks across training, inference, and feedback contexts

## AI+

Identifies system-inherent risks in autonomy levels, capability drift, and emergent behaviors

## Effect+

Captures real-world impact risks on stakeholders, organizations, and society

- Each lens provides a **distinct perspective** on the same AI system
- Together they reveal the **complete risk profile** that traditional AIC classification misses
- Risks **interact across dimensions**—a data quality issue affects model performance, which impacts stakeholder trust
- The framework identifies **blind spots** where organizations have incomplete understanding of their AI risk exposure

# Regulatory Context: EU AI Act

## Risk-based classification is now a regulatory requirement

### EFFECTIVE DATE

**August 2024**

The EU AI Act came into force, establishing the first comprehensive regulatory framework for AI systems

### CLASSIFICATION REQUIREMENT

Organizations must classify AI systems into risk categories: unacceptable, high, limited, or minimal risk

### PENALTIES FOR NON-COMPLIANCE

**€35M or 7%**

Fines up to €35 million or 7% of global annual turnover, whichever is higher

### SYSTEMATIC APPROACH REQUIRED

Organizations must demonstrate systematic risk assessment and classification processes for all AI systems

### The Business Imperative

The 4-Lenses Framework provides a structured method for conducting the risk-based classification that regulators require. It demonstrates due diligence and systematic approach to AI risk management.

# Business Value of Structured Classification

Why organizations adopt the 4-Lenses Framework



## Regulatory Readiness

Systematic approach to risk-based classification required by EU AI Act and similar regulations. Demonstrates due diligence in AI risk management.



## Informed Decision-Making

Clear understanding of AI risk enables targeted governance investments. Resources allocated where they have greatest impact on risk reduction.



## Risk Visibility

Four-lens approach identifies blind spots that traditional frameworks miss. Complete picture of AI risk exposure across all dimensions.



## Stakeholder Confidence

Systematic approach builds confidence among customers, regulators, and stakeholders that AI governance is taken seriously.



## Innovation Enablement

Accurate classification allows appropriate risk-taking for high-value AI initiatives. Move forward with confidence when risks are understood.



## Measurable Returns

30-50% reduction in AI-related incidents through better risk visibility. Faster incident response when risks are well-classified.

# Implementation: Modular and Incremental

Organizations adopt the framework progressively, building capabilities over time

1

## Pilot Assessment

Apply framework to 2-3 representative AI systems to understand classification process and identify gaps

*Timeline: 4-6 weeks*

2

## Inventory Classification

Systematically classify all AI systems in the organization using the four lenses

*Timeline: 2-3 months*

3

## Governance Integration

Embed classification into AI development lifecycle and governance processes

*Timeline: 3-6 months*

4

## Continuous Monitoring

Ongoing risk classification as AI systems evolve and new systems are deployed

*Timeline: Ongoing*

Each phase builds on the previous one. Organizations can **start small** with a pilot and expand progressively as they build classification expertise and governance capabilities.

# Ready to Classify Your AI Systems?

The Enhanced 4-Lenses Framework is ready to use. Start classifying your AI systems today to gain comprehensive risk visibility and regulatory readiness.

1

## Identify Your AI Systems

Create an inventory of AI systems and usage across your organization. Include foundation models, autonomous agents, and AI-powered decision systems.

2

## Apply the Four Lenses

Classify each system through AIC+, Data+, AI+, and Effect+ lenses. Use the framework questions to build comprehensive risk profiles for each AI system.

3

## Prioritize and Act

Use risk profiles to prioritize governance investments. Focus resources on high-risk systems and demonstrate systematic approach to regulators.

[GET STARTED](#)

Contact: **Grio**

[www.grio.nl](http://www.grio.nl)