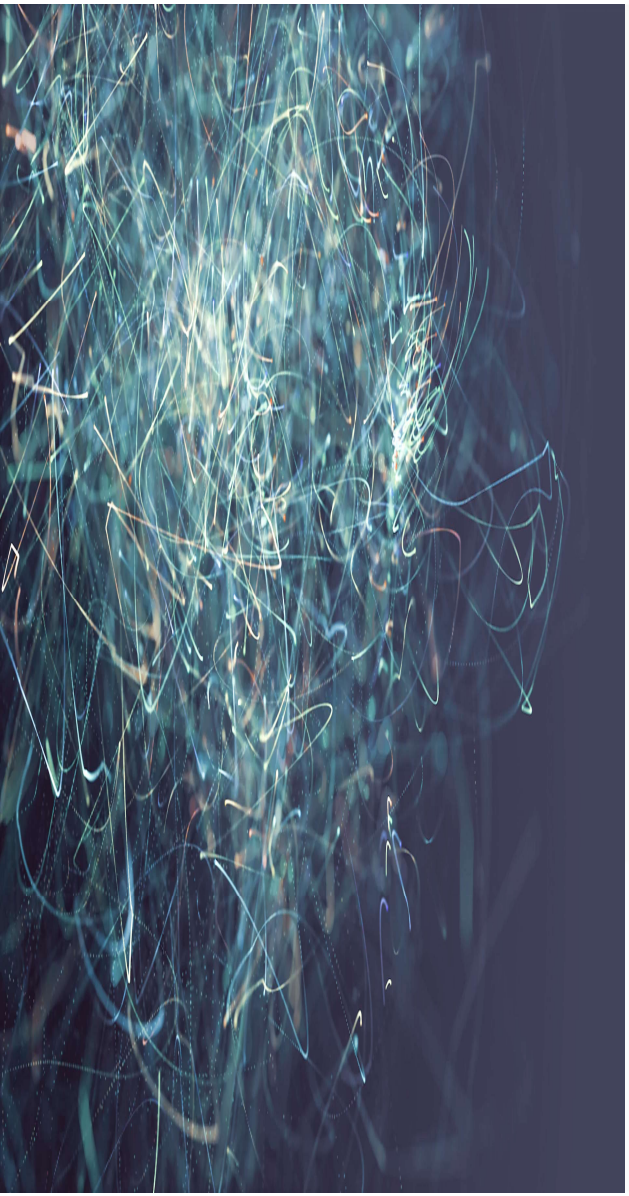
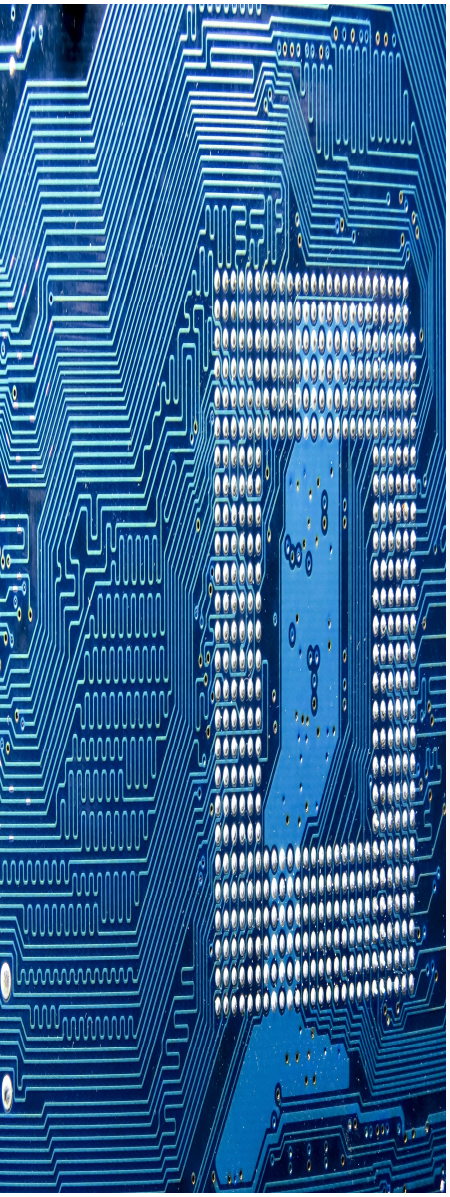


A Practical Guide for Implementing an AI Risk Management Program

Article #1 – The Decide Stage: Should an AI Component be Deployed?

Andrew Sommers, MBA, MS, JD
version 1.0 August 28, 2023

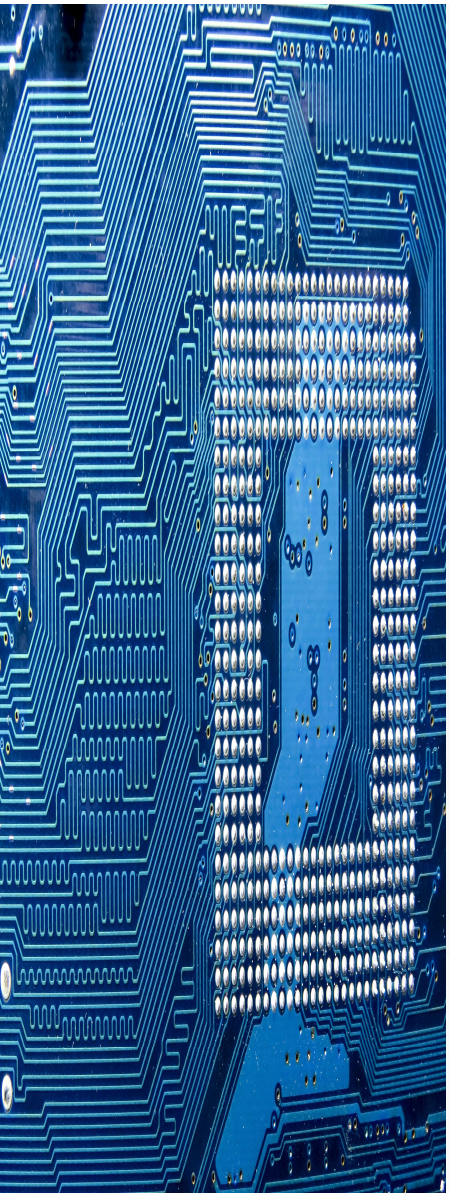


The ‘Decide’ Stage

Prior to deploying an AI component(s) in a system, stakeholders within an organization should review and explicitly agree on the deployment. The initial decision is based on the value provide to the organization from deploying the AI component versus the ability to manage the risks.

The decision is based on the value versus risk decision from the deployment of the specific AI component. A separate process to review the overall project and gain approval to proceed will review the project’s value, costs, and risks. The Decide stage does not include functional, architectural, or operational reviews of the AI component or system. This stage simply reviews the value provided to the organization, the risks from deploying the component, and a high-level assessment of the ability to mitigate unacceptable risks.

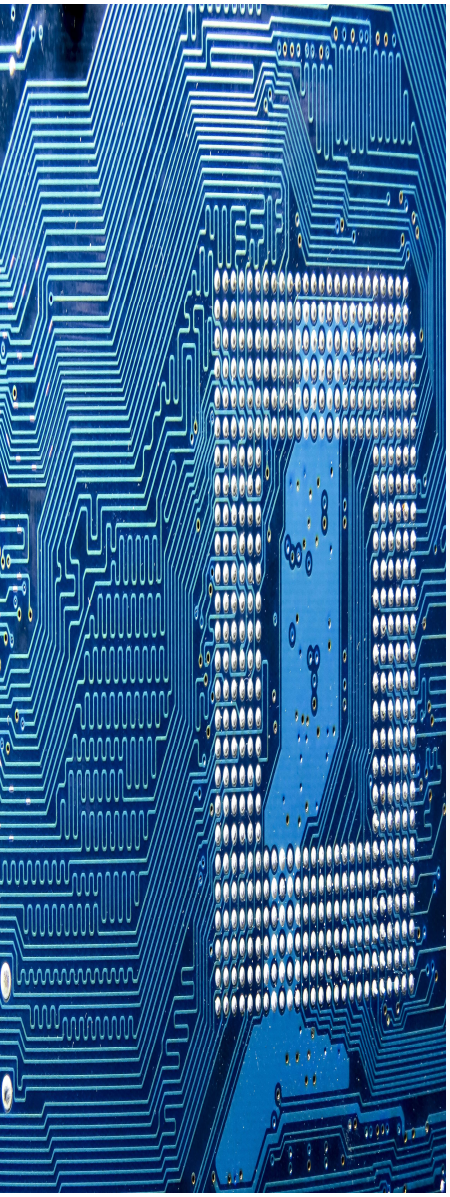
As stated in the Introduction Article for this series, an organization’s exposure to AI will come from multiple sources including internal development, custom solutions developed by external parties, vendor provided systems, and reliance on third party service providers who make use of AI in their systems. The Decide stage should be applied to all sources of exposure to AI components.



The Process for the Decide Stage

1. Determine if a component should be classified as AI.
2. Determine the value provided by the AI component.
3. Discuss the risks and ability to mitigate the risks from deploying the AI component.
4. Determine the need for an enterprise level review.
5. Finalize the decision.

This article will discuss each step in the process and provide examples of decision frameworks.



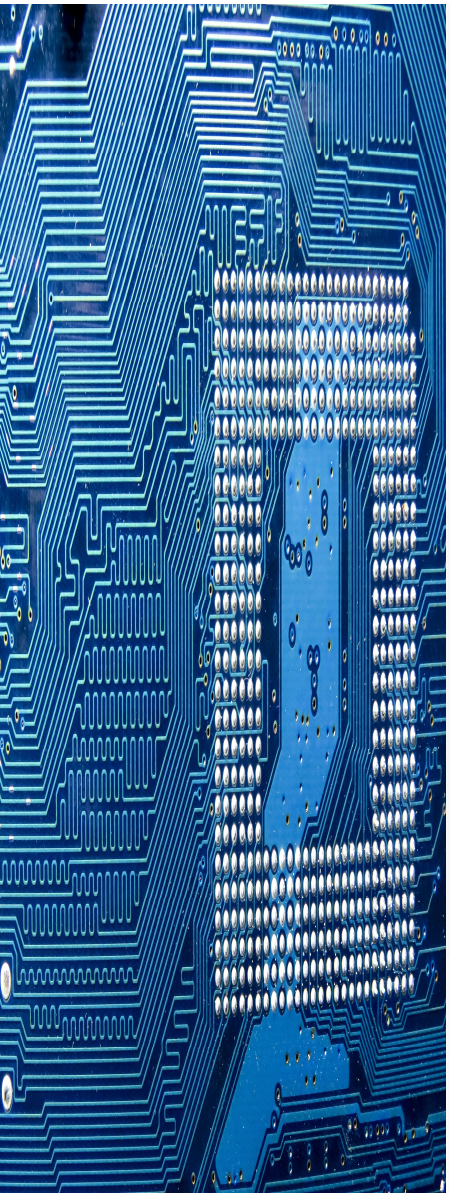
Who Are the Decision Makers?

Before developing each step in Decide stage, determine who should make the decision to deploy an AI component.

An organization does not have the resource capacity to perform an enterprise level review of every potential deployment of an AI component. Only higher impact AI component deployments should be reviewed at the enterprise level. Lower impact deployments should be reviewed by the team planning to implement the system containing the AI component. The discussion of Step 4 in the decision process, presented later in this article, provides guidance for a team to determine the need to elevate a deployment decision to the enterprise level. For all decision steps in the process, start the decision process at the team level and elevate the decision to the enterprise level as required.

At the team level, decision makers should have a thorough understanding of the team's strategy and operations. The enterprise decision makers should be a group of leaders who can represent the organization and should include representatives from enterprise functions including legal, compliance, risk management, and data governance. At both levels, the core decision making group contains a limited number of standing members. Specialists are added to the decision group based on the characteristics of the AI component and the system in which the component will be deployed. Smaller decision groups will be more efficient at decision process, but specific expertise will be need to inform the decision makers.

The decision framework examples provided in this article can be customized by an organization based on its potential deployment of AI components. The resulting decision frameworks should be applicable for both team and enterprise level decision making.



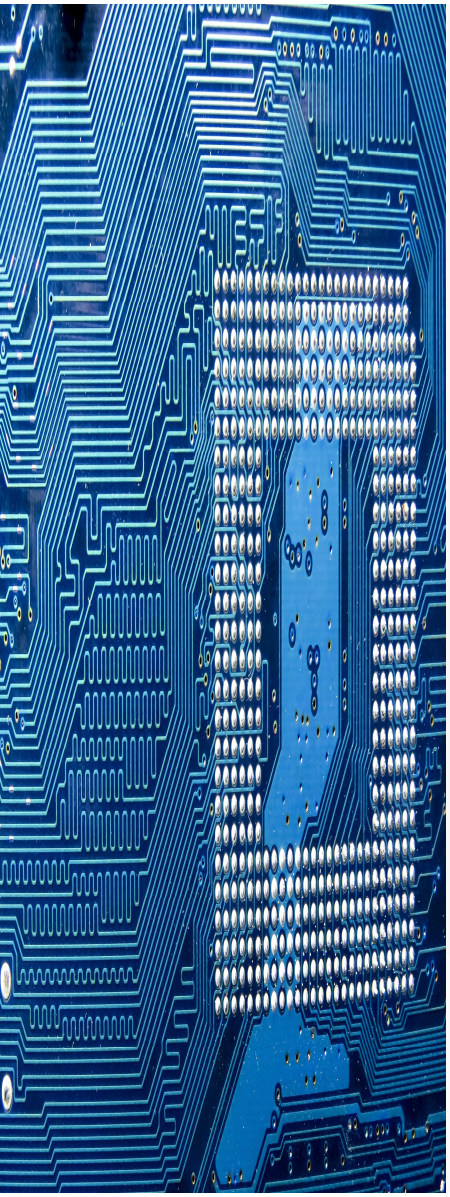
1. Determine if a Component Should be Classified as AI

As noted in the Introduction article, system stakeholders will disagree on the classification of a component as AI. The European Union's Artificial Intelligence Act[1] ('EU AI Act') defines AI in general terms, but an organization requires a more prescriptive approach to classifying components.

The EU AI Act provides more specific guidance for identifying a component as AI in Annex I by listing the following methods to be consider as AI:

'(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods.'

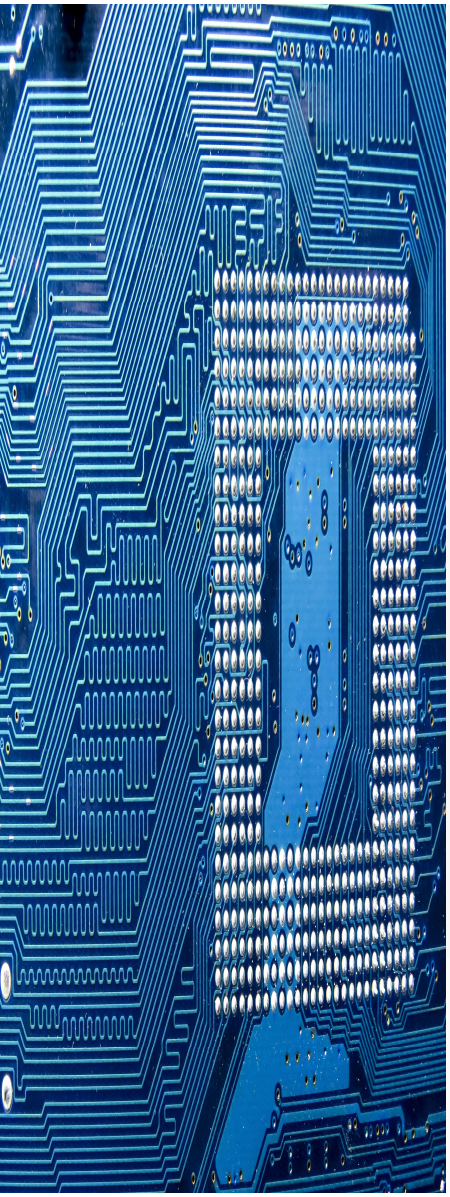
The additional guidance is more prescriptive and technical; This guidance is likely not sufficient for an organization to determine if a component should be classified as AI. An organization should build a decision framework to determine if a component should be classified as AI based on functionality factors. An example decision framework is provided in Appendix I. The example can be used as a starting point to customize a decision framework for the organization.



2. Determine the Value Provided by the AI Component

The determination of the value provided by deploying an AI component focuses on the value added by the component rather than the value of deploying the entire system. Appendix II provides an example framework to determine the value added by deploying the AI component. This example is based on the Balance Scorecard[2]. The Balance Scorecard was developed to align individual undertakings within an organization to enterprise level objectives; The framework can be modified to support the decision to deploy an AI component.

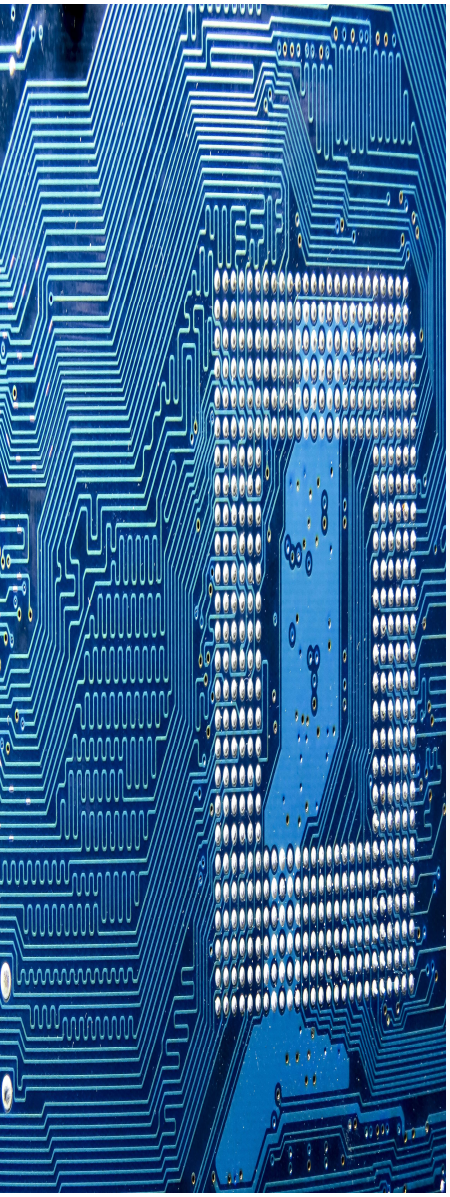
As with other decision examples provided in this article, leveraging an existing and standardized framework as the basis for a decision model allows an organization to leverage the background, guidance, and continued development of that framework.



3. Discuss the Risks and Ability to Mitigate the Risks from Deploying the AI Component

Deployment and continued use of an AI component generates risks. Some of the risks might not exist at initial deployment but manifest during continued operation. All foreseeable risks should be considered when deciding to deploy an AI component. The risk analysis framework identifies the risks, provides a measure of the likely impact, and provides a means to discuss the high-level approach to addressing any risk considered unacceptable by an organization. An AI component should not be deployed if the organization does not have confidence that an unacceptable risk can be sufficiently mitigate, or if the value proposition from Step 2 is insufficient given the risk profile of the AI component.

The example risk assessment framework in Appendix III is based on the OECD's Framework for the Classification of AI Systems[3]. The OECD Framework was developed to classify AI systems to support policy development. With modification, the framework can be used to assess risk and discuss risk mitigation of an AI component. The OECD provides rich context to support the Framework and will continue to develop this tool. Leveraging the OECD Framework to develop a risk analysis framework for deployment decisions allows an organization to leverage OECD guidance.



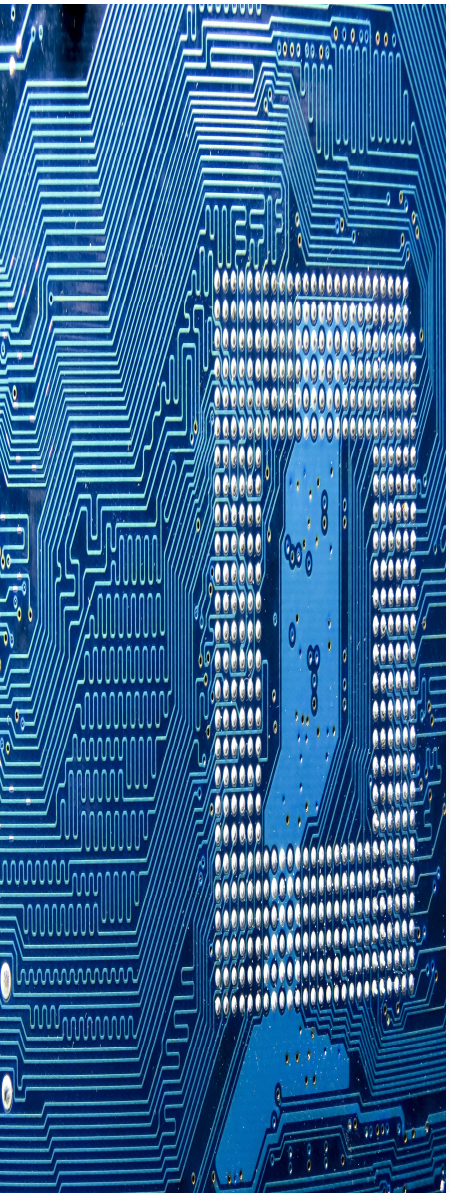
4. Determine the Need for an Enterprise Level Review

The initial development an analysis of the value and risks of the AI component will start at the team level. To maximize resources and support agile decision making, decision making for deployment of an AI component should occur at the team level unless the team determines an enterprise decision is required.

Consider elevating the decision to deploy and AI component to the enterprise level decision group if the system will impact:

- Delivery of products or services to customers;
- Relationships with critical business partners;
- Employee recruitment, development, or opportunities;
- Mission critical operations;
- Regulatory compliance; or
- Cross-functional operations.

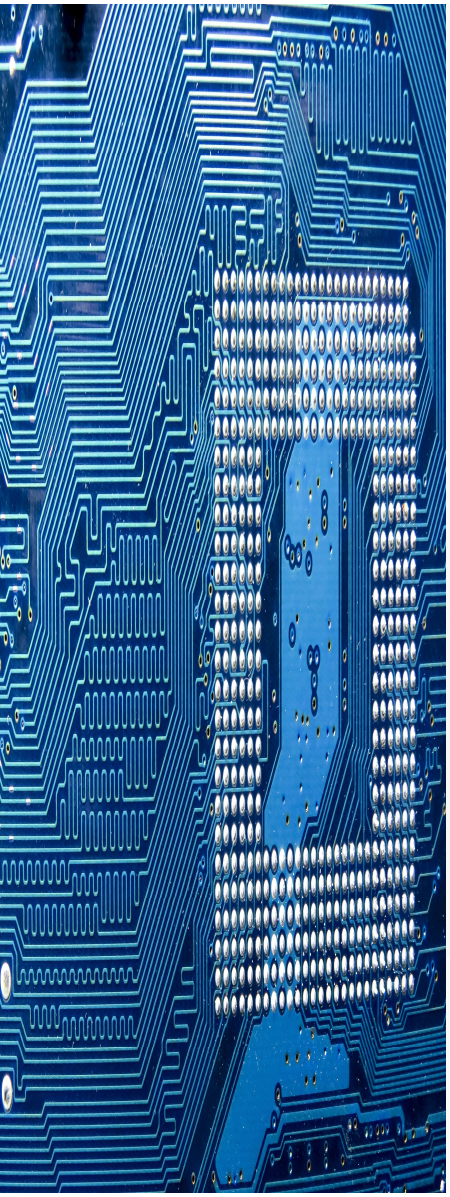
To support enterprise transparency to the deployment of AI components in an organization, all team level decisions to deploy AI components should be socialized with the enterprise level decision group and documented as discussed in Step 5.



5. Finalize the Decision

Based on the analysis of the value and the determination of the acceptability of the risks to achieve the value, the decision group will either reject or approve the deployment of the AI component. Based on the on-going or emergent risks from continued use of the component, the decision group might schedule follow-up review(s).

Again, any team level decision to deploy an AI component should be socialized with the enterprise level decision group. The decision and a brief description of the component and system implementing the component should be documented as part of maintaining an enterprise inventory of deployed AI components. The documentation to support the decision should be maintained in the enterprise inventory. Documentation should be updated based on continued reviews.



Limitation of Scope

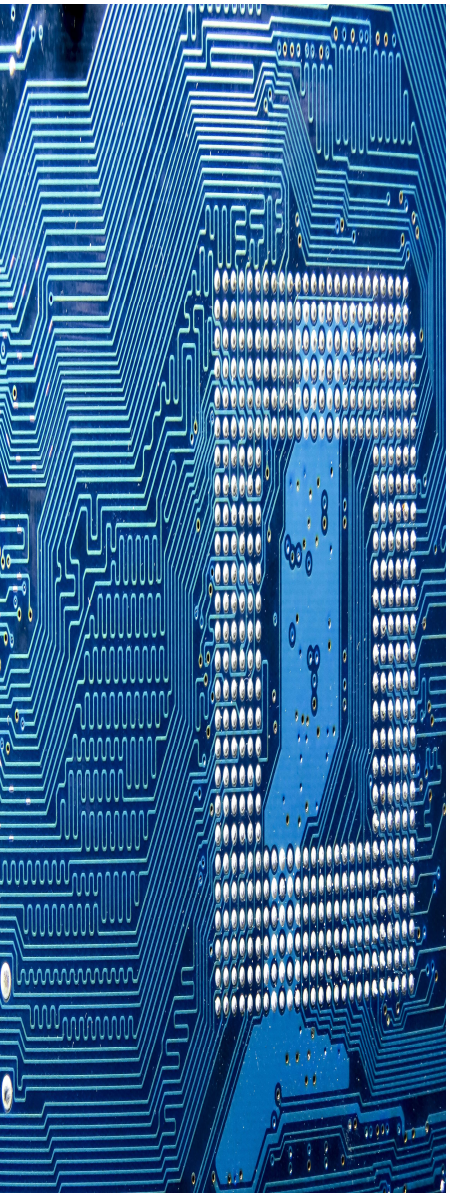
The series will discuss where the AI risk management program should ensure alignment with the company's data governance, security, and privacy programs. While data governance, security, and privacy are vital parts of managing the risk of systems that include an AI component, this series will not provide detailed guidance on these risk management functions. Other sources should be consulted for details of implementing these data and risk management functions.

Notes

The series is not and should not be used as legal advice or recommendations.

The articles provide practical guidance on developing an AI risk management program but should not be viewed as a comprehensive guide to AI risk and risk management and should be supplemented with other sources.

Articles might be updated from time to time to reflect updates to AI components, regulations, and risk management frameworks. See the version number and date on the initial page of this document.



References

1 European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts, April 21, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

2 The Balanced Scorecard – Measures That Drive Performance, Robert S. Kaplan and David P. Norton, Harvard Business Review, January 1992. https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2

3 OECD Framework for the Classification of AI Systems, OECD Digital Economy Papers, OECDPublishing, No. 323, February, 2022. https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm

Images provided by pxfuel at https://www.pxfuel.com/

Appendix I – Decision Framework to Determine if a Component Should be Classified as AI

This framework uses functional factors to determine if a component should be classified as AI. If a 'Yes' appears in all three sections, the component should be considered AI and proceed through the appropriate review process.

Decision Criteria	Yes / No
1. Does the component perform one of these tasks?	
Classify a person, item, or observation into a category	
Identify a specific person or item	
Recommend or take an action	
Generate a recommended output or outcome	
2. Does the component learn from examples?	
Supervised (using labeled examples or human guidance)	
Unsupervised (using unlabeled examples)	
Through reinforcement (adjusted based on evaluation of outcomes)	
3. Does the deployed component generate its output without on-going human intervention?	

Question 3 addresses the use of a models as an analysis tool versus a decision component. For example, a simple distribution analysis to assist an individual with determining the likelihood of an outcome would not constitute an AI component in a system. However, a component that develops statistical distributions based on examples and generates recommendations or output without on-going human intervention would likely qualify as an AI component under the EU AI Act.

Appendix II – Decision Framework to Determine the Value from Deploying an AI Component

Four dimensions are represented in this framework, Customer, People, Internal Process, and Financial Impact. While the Financial Impact is not explicitly defined as a dimension in the framework, outcomes such as 'Increase in Sales', and 'Lower Costs' provide an opportunity to consider the financial impact of deploying an AI component. The input in the three right hand columns can be values, value ranges, or qualitative assessments.

Customer	Increase in Sales	Improve Relationships	Improve Customer Retention
New or improved product			
Better customer service			
Improved customer contact			
Broader product distribution			
People – Recruitment, Learning, and Growth	**Increase in Diversity**	**Improve Productivity**	**Improve Employee Retention**
Broader or more diverse candidate reach			
Better alignment of job opportunities with skill sets and interests			
Improvements in learning and development			
Internal Process	**Improve Delivery**	**Lower Costs**	**Reduced Errors**
Automation of manual tasks			
Improvements of the expertise within a system or process			
Increased monitoring of system or process tolerances			

Appendix III – Decision Framework to Assess the Risks from Deploying an AI Component

This framework can be adjusted to measure the likelihood and impact of a risk following the high-level approach to mitigating any unacceptable risks (add columns to the right). Risk mitigation requires only a high-level assessment for this analysis.

Impact to	Risk	Likelihood	Impact	Risk Score	Approach to Risk Mitigation
System User Interaction					
User Opt-out	System user cannot opt-out of the AI based interaction that is not providing a useful result				
User Desertion	Poor quality of user interactions leads to session abandonment and limited repeat usage				
Personal Rights					
Privacy Violations	Personally identified data is used in violation of privacy regulations				
Impact to Fundamental Rights	Decisions regarding employment, education, or other fundamental rights are unfair				
Ability for Human Intervention	A mechanism does not exist for a person to address an individual's concern or complaint about possible unfair outcomes				
Data					
Availability of Data	Sufficient data to represent necessary observations is not available				
Data Rights	The required rights to use the data with the system cannot or are not obtained				
Historical Bias	Data contains a historical bias or gap that cannot be addressed with supplemental data				
On-going Data Availability	Updated data to support continued use of the AI components is not available				
AI Model					
Transparency	System owners do not understand the role of an AI component(s) in the system				
Explainability	Outcomes of the AI components cannot be explained or attributed to specific components				
Continued Learning	Continued learning by the model leads to undesirable outcomes				
Operations					
Skill Set	Skills required to operate the system are not readily available				
Optional Dependency	Critical business function(s) will not be operational if systems fails				
Operator Override	The system cannot be overridden when outcomes are determined to be suboptimal				