RISK MANAGEMENT FRAMEWORK

for the

Procurement of AI Systems™

(RMF PAIS 1.0)

Dr. Cari L. Miller, cari@inclusivechange.org
Gisele Waters, Ph.D, partner@innovationresearch.com

AI Procurement LAB

The Center for Inclusive Change

Abstract

Classifying and understanding risk is essential when procuring an AI system. It is important to recognize that the types and scale of risks vary from system to system. When AI systems are developed for high-risk domains (e.g., employment, health, education, housing, finance, public assistance, etc.), two risk indicators become highly relevant. These indicators include 1) the complexities within the AI system and 2) the impact that outcome(s) may have on human lives. Hence, it is imperative to determine how much risk the procuring organization is willing to accept for each system at the outset of each procurement. This act is known as establishing the risk appetite for the procurement. A well-defined risk appetite for a procurement should serve as an anchoring point throughout the procurement lifecycle to guide risk identification, risk treatment, risk controls, and risk monitoring strategies to create an acceptable risk tolerance for the chosen system—in order to enjoy the AI system's intended benefits more fully.

Keywords—procurement, acquisition, public service, artificial intelligence, responsible AI, high-risk AI, risk management, NIST AI RMF

Table of Contents

1.0 Introduction

Continuous improvement of operational efficiencies and decision-making effectiveness is a common and customary organizational objective.[1] In addition, organizations have a basic but important requirement to uphold jurisdictional laws, regulations, and a duty of care for all stakeholders.[2] These types of organizational objectives demand that any tool, in this case an AI system,[3] must not only deliver on its promises to improve organizational circumstances (such as operational efficiency and decision-making effectiveness), but it must also operate within the organization's acceptable risk appetite.[4]

Unfortunately, certain AI use cases can present new and novel risks to organizations (beyond reputational and legal damages). Such use cases include systems designed to deliver critical, sometimes life-altering, decisions in the context of employment, health, education, housing, finance, public assistance, critical infrastructure, essential utilities, law enforcement, immigration, justice, legal services, biometric identification, safety components and other consequential decision systems. Systems developed for these types of use cases are commonly referred to as "high-risk" systems.[5] They pose a high-risk to those impacted either directly or indirectly by the AI system in question in significant or critical ways.

The Risk Management Framework for Procuring of AI Systems (RMF PAIS 1.0), discussed in Section 4.0, focuses primarily on risk management for high-risk systems. High risk systems are further defined in Appendix A. More specifically, given that high-risk systems can produce great advantages in terms of efficiency gains and consistency in decision-making output,[6] they also have the potential to impact a person's safety, civil rights, and/or fundamental human rights and dignity.[7] AI researchers have repeatedly identified a variety of harms that require our attention when deploying such systems in general and especially when deploying high-risk AI systems.[8] The known harms that have been perpetrated by these systems are particularly troubling because they stem from historic systemic bias that can be widely scaled across vulnerable populations through

[1] Azeem, M., Ahmed, M., Haider, S, & Sajjad, M. (2021). Expanding competitive advantage through organizational culture, knowledge sharing and organizational innovation. *Technology in Society, 6*(101635). https://doi.org/10.1016/j.techsoc.2021.101635

[2] https://www.financialexecutives.org/FEI-Daily/March-2022/Duty-of-Care-The-Board%E2%80%99s-Role-in-Navigating-Fores.aspx

[3] The definition of AI systems is based on the OECD's definition of AI systems. https://oecd.ai/en/wonk/ai-system-definition-update

[4] Rittenberg, D. L. & Martens, F. (2012). *Understanding and communicating risk appetite*. The Committee of Sponsoring Organizations of the Treadway Commission. https://www.coso.org/Documents/ERM-Understanding-and-Communicating-Risk-Appetite.pdf

[5] https://competitionlawblog.kluwercompetitionlaw.com/2023/06/02/deployers-of-high-risk-ai-systems-what-will-be-your-obligations-under-the-eu-ai-act/

[6] Schmarzo, B. (2023). AI & data literacy: Empowering citizens of data science. Packt Publishing.

[7] Office of Management and Budget. (2023, October 30). *Proposed memorandum for the heads of executive departments and agencies*. https://ai.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-Public-Comment.pdf

[8] https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf https://www.holisticai.com/blog/whitepaper-us-ai-regulation

powerful and non-transparent algorithmic computations,[9] generating unfair, unequal, disproportionate, and potentially life-altering outcomes.[10]

Without proper governance and risk mitigation practices, the computational power of high-risk systems can pose unwanted threats to our fellow humans--causing more than just reputation or legal damages for organizations. Unmitigated risks and negative outcomes can, in some cases, mean life or death to the end uses. Hence, the stakes are high, and organizations must find ways to control the risks in order to establish trust for all stakeholders.[11] The RMF PAIS 1.0 provides a guide to identifying and controlling these risks through the use of standard procurement lifecycle processes in a practical and responsible way.

Assuming an organization has identified a legitimate problem or business need for which the only solution is to procure an AI system, the first step in the framework requires a team of individuals to develop the risk appetite for the procurement. This critical step sets the risk ceiling that is applied throughout the balance of the procurement lifecycle. Once the risk appetite is established, the procurement team can define the requirements for the solution based on the contours of the risk appetite. Vendors will then utilize the solution requirements to develop their proposals. Subsequently, the vendor proposals should be assessed against the parameters of the risk appetite to identify any risk exposure evident in the proposed systems that may exceed the risk appetite. Ultimately, risk mitigation tactics and risk tolerance metrics should be negotiated with the vendor(s) that best meet the solution requirements and risk profile. The agreed upon mitigation tactics and risk tolerance metrics should be incorporated into the procurement contract as risk control mechanisms with clearly articulated accountabilities. Those risk controls and tolerance should then be used to conduct ongoing monitoring and management of the system to ensure that any realized risks do not exceed the predetermined risk appetite.

2.0 Framing

The RMF PAIS 1.0 is specifically designed for *buyers* of AI systems. It is not meant to replace current procurement practices, but rather augment and strengthen those practices. Furthermore, the framework is a traditional risk management framework (RMF) modeled after ISO 31000, COSO, and ISO 42001 (noted below) with adjustments, modifications, and adaptations made to address risks related to human-sensitive aspects (e.g., civil rights, human rights, dignity, etc.) of socio-technical systems that are not present in RMFs designed to address financial and enterprise assets (e.g., new facility investment, power feed redundancy investment, new phone system selection, etc.) decisions.

[9] O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. United Kingdom: Crown.

[10] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. United States: St. Martin's Publishing Group.

[11] European Commission. (2020). *The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment*. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

2.1 Buyers, not Developers

The AI ecosystem comprises many stakeholders including end users/data subjects, administrative users, buyers/deployers, developers, AI supply chain "component providers" (including data brokers and open-source model developers), government legislators/policy makers/regulators, the judicial system, and society at large. Much of the AI governance literature tends to focus on steps that AI *developers* should take to ensure that their systems are responsibly and ethically designed.[12] Given that AI developers hold the largest amount of authority over deciding the risk mitigations that are designed into their systems, this makes sense.

That said, this framework is written for use by AI system ***buyers***, not developers. Buyers of AI systems do not have control over the design decisions and choices made by developers. As such, it is the buyer's responsibility to verify that those decisions and choices were conducted in a manner that is appropriate and befitting of the buyer's risk appetite.

2.2 Intersection with Current Procurement Practices

This framework is not meant to replace any aspects of current procurement practices. Rather, it is meant to augment and strengthen those practices. While current procurement practices will uncover valuable benefits of AI systems, they are less likely to uncover the unique risks that AI systems present (beyond the risks that are common in traditional IT systems). Hence, the RMF PAIS 1.0 does not duplicate common procurement practices. The emphasis here is on the risk-based aspects of the AI system that organizations must identify, treat, control, and monitor; All done so that the risks do not overwhelm the benefits and intended return on investment discovered through existing and traditional procurement practices.

2.3 Guiding Risk Management Frameworks

Risk management was originally born within the financial sector in the late 40's and early 50's, but began to flourish intensely in the 70's.[13] Over time, a broader view of enterprise risk management has evolved to include also legal risks, supply chain risks, operational risks such as compliance, fraud, employee retention, IT system disruptions, and more.[14, 15] In order to make risk management a predictable organizational practice, risk management frameworks (RMFs) were designed to guide teams through risk management process steps. Traditional RMFs such as ISO 31000 – Risk Management Standard,[16] COSO enterprise risk management – integrated framework,[17] and ISO 42001 – Artificial Intelligence Management Systems [18] (which are the models for the RMF PAIS 1.0) demonstrate a consistent pattern of process steps including:

[12] IEEE. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, Version 2 for public discussion. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

[13] Dionne, G. (2013). Risk management: History, definition, and critique. *Risk Management and Insurance Review, 16*(2), 147-166. https://doi.org/10.1111/rmir.12016

[14] See footnote 12.

[15] Nazarov, M. (2023, August 8). *Enterprise risk management (ERM) fundamentals* AuditBoard. https://www.auditboard.com/blog/enterprise-risk-management/

[16] https://www.iso.org/iso-31000-risk-management.html

[17] See Footnote 4.

[18] https://www.iso.org/standard/81230.html

- Leadership commitment (policies, roles, and responsibilities)
- Risk appetite (amount and type of risk)
- Risk-aware requirements (internal and external needs)
- Risk assessment (identify, prioritize, mitigation mapping)
- Risk control (avoid, accept, mitigate/transfer/share)
- Risk monitoring (survival, measure, review, triage)

To be clear, the RMF PAIS 1.0 was carefully aligned to a traditional RMF to maximize its implementation, adoption, and effectiveness, which is to say that it goes beyond other AI procurement guides and frameworks (e.g., World Economic Forum's Procurement in a Box, UK's Guidelines for AI Procurement, U.S. Government Accountability Office AI Accountability Framework for Federal Agencies) that contain many risk-based questions without also providing risk management process discipline. However, while the RMF PAIS 1.0 is modeled after ISO 31000 and COSO, it deviates slightly in order to address (and respect) risks that are specific to *human* impacts of socio-technical AI systems where traditional RMFs more often focus on financial/shareholder protections and/or risks related to physical continuity of business operations (e.g., plants, facilities, equipment, etc.).

3.0 Prerequisites

The RMF PAIS 1.0 is designed to address risks directly related to the procurement at hand. That said, there are several important prerequisites that are needed to properly inform the RMF process. While it is widely understood that general AI literacy is necessary in the 21st century, conducting AI procurements responsibly requires some manner of *advanced* AI literacy and legal/policy literacy. In other words, knowledge of certain technical and managerial decisions within the AI lifecycle are necessary and informative to the risk identification and risk management process. Additionally, legal and policy issues are self-perpetuating and must be continually monitored for obvious reasons of compliance. Beyond the literacy prerequisites, procurement professionals should also maintain a clear understanding of their organization's data quality and user AI literacy levels. Each of these aspects of responsible AI procurement also has a direct impact on establishing sound risk management tactics.

3.1 AI Acquisition Literacy

General AI Literacy. A basic understanding of AI concepts[19] such as recognizing diverse types of AI, understanding how AI can impact individuals, and knowing how users and or subjects of AI systems can (and should) safeguard themselves.

Advanced AI Literacy. Deeper knowledge of AI concepts especially as they relate to the procurement of AI. While acquisition professionals do not need to be coders and technical engineers, they do need to understand the AI lifecycle, technical elements of an AI systems (e.g., data, models, features, and functions, etc.), and the more nuanced extraneous aspects of AI systems

[19] Long, D., & Magerko, B. (2020). *What is AI literacy? Competencies and design considerations*. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376727

that impact its performance and outcomes (e.g., ethical choice, development procedures, the AI value chain, testing, explainability, etc.). Understanding these technical and managerial factors is critical to effective risk identification and management practices. The RMF PAIS 1.0 assumes that the user of the framework understands these more advanced elements in AI literacy.

Legal and Policy Literacy. In addition to the above, AI laws, regulations, and even organizational policies continue to evolve at a rapid pace. The RMF PAIS 1.0 assumes that the users have a clear understanding of the relevant laws, regulations, and policies that may impact each procurement. For example, the EU AI Act of 2024 outlines several types of AI systems that are prohibited.[20] The framework assumes your organization knows what is or is not allowed from a legal and policy perspective. In other jurisdictions, laws are being enacted to address AI in specific domains or for specific use cases like insurance, human resources, warehouse work, fair housing, etc. In still other cases, laws and policies are being set out to address specific types of AI such as facial recognition, biometric data, deep fakes, LLM content provenance. Staying abreast of these developments is an essential aspect of ongoing risk management.

3.2 Organizational Readiness

Procuring Enterprise: AI Policies, Roles, And Responsibilities. The RMF PAIS 1.0 can serve as a guide for procurement teams to address AI governance matters that are directly related to the AI system being procured such as specific risk treatments and controls associated with the system and the outcomes. However, it is up to the procuring organization to establish organizational-level AI governance that addresses the good stewardship of their own resources. For example, organizational AI governance mechanisms and tactics should include clear roles and responsibilities that oversee and are accountable for responsible AI procurement and deployment, ongoing monitoring, and AI incidents as well as establishing and maintaining organizational policies, practices, and procedures for effective and consistent administration of communications, notices, and compliance requirements (e.g., transparency, interpretability, and explainability), incident management procedures, timely adjudication and redress of AI incidents, whistleblowing procedures, and administrative user and end user education and training to name a few areas of organizational governance concern.

Procuring Enterprise: Data Maturity. Data is like gasoline for AI systems. When the data is not optimized and/or is not fit for the intended purpose of the system, the expected performance will be compromised and the risks for usage will skyrocket. In an ideal world, all data would be perfectly optimized and always fit for the correct purpose. However, that is not a world we live in. As such, it is important for procurement professionals to understand where their internal data maturity levels sit. Given that data will range from department to department and silo to silo, it should be expected that the quality of data will vary across the organization. Hence, prior to embarking on any procurement that may look to internal data to "feed" the AI system, assessing the maturity level of the input data will be an important prerequisite piece of knowledge to carry into the procurement process.

[20] https://www.euaiact.com/article/5

Procuring Enterprise: User AI Literacy Maturity. Just as it is important to understand the maturity levels of data within the enterprise, it is equally important to understand the maturity levels of AI literacy among the enterprise workforce and/or external users that will interact with the intended AI system. For example, if the users are IT engineers that build machine learning systems, it may be safe to assume that their AI literacy levels are greater than users that are call center operators having high school degrees. These are key factors when considering that even simple AI systems can present operational risks (e.g., propensity for misuse) when deployed to users with low AI literacy. Hence, an important prerequisite for every procurement professional is to know who will ultimately interact with the AI system and how. This will help contextualize risk judgments when assessing, identifying, treating, controlling, and monitoring risks that a chosen AI system could present.

3.3 Legitimate Business Need

A thorough root cause analysis of the business need (e.g., a business problem or an opportunity) is necessary to create successful risk management strategies and tactics. While a full discussion of how to define a business need through root cause analysis is out of scope for this paper, more information on the topic is readily available elsewhere online.[21]

Most importantly, the results of a root cause analysis should provide critical background information to the depth and complexity of the potential risks within the problem or opportunity. The types of information that are relevant to risk management include but are not limited to the duration, timing, and frequency of occurrences; the types and quantities of stakeholders involved; the outcomes and impacts each stakeholder group has been experiencing; the nature, availability, and quality of the data involved in the decision processes; the ability, timing and frequency of the stakeholders requiring and receiving redress for unfair outcomes; and the organization's readiness to govern and manage an alternative system. Understanding the contributing elements of each problem or opportunity for *each* procurement is an important lens through which the risk management framework should be applied.

4.0 RMF PAIS 1.0 Overview

The RMF PAIS 1.0 (see Figure 1 below) contains five steps that align with traditional risk management framework constructs[22] tailored to risks related to AI systems that can be managed across the procurement lifecycle. In addition, the RMF PAIS 1.0 also aligns with the National Institute of Science and Technology AI Risk Management Framework (NIST AI RMF 100-1)[23] and several other AI-oriented and sector-based risk management frameworks and instruments, which are further discussed below.

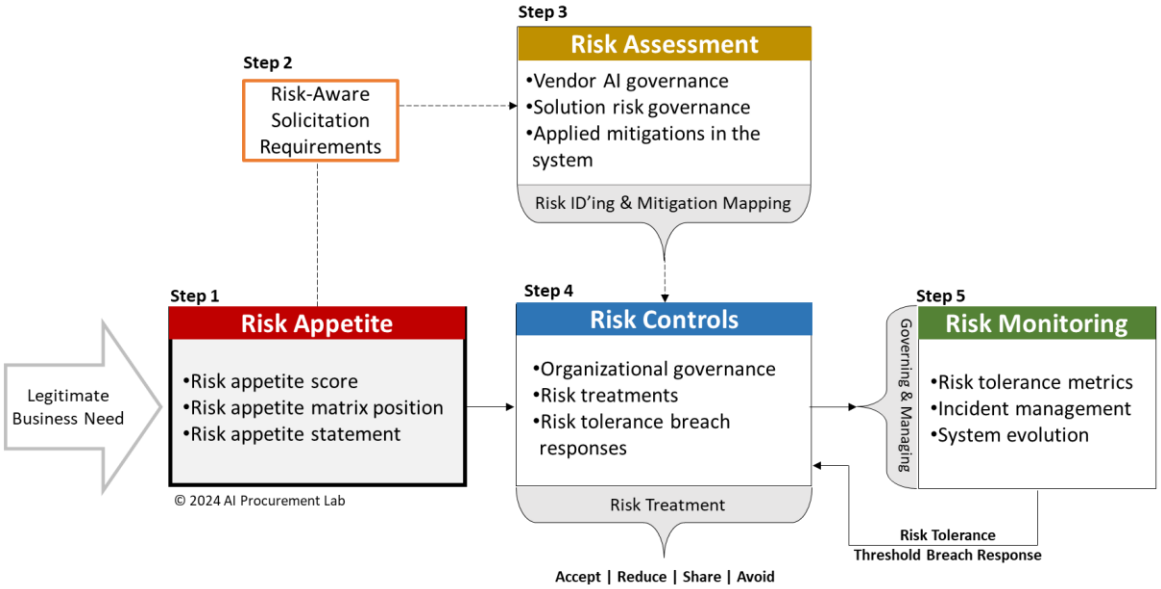
[21] https://thecompassforsbc.org/how-to-guide/how-conduct-root-cause-analysis

[22] See Footnote 16.

[23] National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. https://doi.org/10.6028/NIST.AI.100-1

Figure 1.

Risk Management Framework for Procuring AI Systems



5.0 STEP 1: Risk Appetite

All organizations desire to improve through many means. More recently, organizations are increasingly turning to AI systems to facilitate their improvement goals and objectives. However, while AI is perceived to provide efficiency gains and decision-making improvements among other benefits for organizations, it is common for these benefits to coexist with known and unknown harms comingled within the same AI systems. That said, no two AI systems are alike. *Each AI system will contain unique circumstances, stakeholders, stakeholder dimensions, input/output relationship complexities, intended outcomes, potential benefits, and possible harmful impacts.* Thus, each procurement will have a unique risk profile, which means the procurement team must determine how much risk the organization is willing to "accept" from the vendor(s) they select to deliver the AI system. The amount of "acceptable" risk is called the risk appetite.

The primary purpose of a risk appetite is to provide a necessary guidepost for the procurement team to recognize vendors, systems, and contract terms that exceed the organization's risk tolerance for a particular system.[24, 25] As a secondary purpose, determining the risk appetite up front can also mitigate unfair biases towards known or preferred vendors that may arise from human reviewers.

As such, the first step in the risk management framework of AI procurement entails setting a risk appetite, which is then used as a critical lens to use during the procurement lifecycle when

[24] Casovan, A. & Shankar, V. (2022). A risk-based approach to AI procurement. *The Legal Review*. https://www.theregreview.org/2022/07/11/casovan-shankar-a-risk-based-approach-to-ai-procurement/
[25] See footnote 14.

acquiring an AI system.[26] Done correctly, establishing a risk appetite for each procurement will directly relate to the desired outcomes the organization is seeking while concurrently managing the associated risks. For example, while an AI system can improve efficiency, the possibility of discriminatory output from the system could also pose serious legal consequences. As such, establishing an appropriate risk appetite at the beginning of the procurement lifecycle will guide the team's assessment, negotiations, and decision-making to ensure risks are identified and controlled in order to set tolerable risk levels that meet organizational goals, such as establishing inclusivity and avoiding discriminatory acts.

5.1 Differences between AI Procurements

Exhibits A and B in the adjacent text box represent two distinct types and magnitudes of risks.

In Exhibit A, the risks are great. Many people are involved. Highly complex and technical medical data points are involved to determine the decision, and the algorithms in the system include a complex neural network. Thus, the AI system is less transparent and more challenging to explain. Further, the algorithmic output has a direct impact on vulnerable individuals' ability to feed, clothe, and shelter themselves and their families.

Exhibit B, on the other hand, has a much smaller target audience. The AI driving this system is based on less complex algorithms. Theoretically, the output does not pose a threat to a human's ability to feed, cloth, and shelter themselves or their family.

EXHIBIT A: An AI system that uses 150+ data points to determine the financial benefits that millions of people will receive for their disabilities each month.

EXHIBIT B: An AI chatbot on a high school website that helps parents understand the school's student behavior policies.

As a result, these two systems have very different risk profiles. The system described in Exhibit A is high risk. The system described in Exhibit B may be deemed low to medium risk. The key takeaway here is that understanding the purpose, population, and parameters involved in each AI system is essential for the procurement team to establish a relevant risk appetite for each procurement.

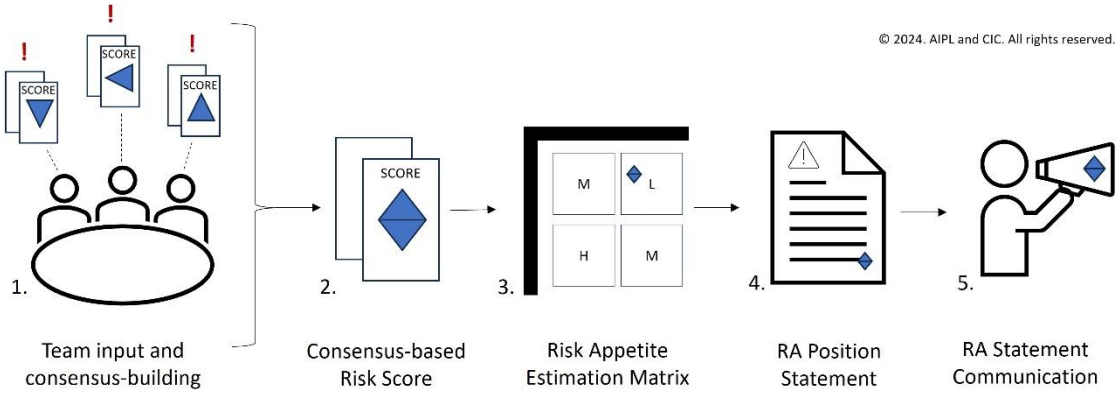
5.2 Determining the Risk Appetite

There are five steps in determining, evaluating, and establishing an organizational consensus-based risk appetite. (See Figure 2 below.) These first steps involve three key elements including a risk appetite (RA) scorecard, an RA matrix, and several corresponding RA risk statements. Each will be explained in more detail in the sections below. This section will provide a high-level overview of the five steps in setting a risk appetite.

[26] Risk Leadership Network. (2020). *What is your risk appetite and how do you implement it?* https://www.riskleadershipnetwork.com/insights/what-is-risk-appetite-and-how-do-you-implement-it

Figure 2.

Risk Appetite Process Steps



- Step 1: The team[27] most familiar with the requirements of the procurement will each fill out a risk appetite scorecard (Appendix B). The scorecards consist of two dimensions of risk discussed in detail below.
- Step 2: Once all scorecards are completed, the team will work together to negotiate a single scorecard that represents a consensus based on their collective inputs.
- Step 3: The team will use the risk appetite matrix and plot the two scores from the RA scorecard into the 2x2 RA matrix. This will indicate the risk appetite for the procurement at hand.
- Step 4: The team will choose the corresponding risk appetite statement to help convey the level of risk tolerance the organization is willing to accept throughout the procurement at hand. In some cases, the team may decide to adjust and tailor the language to match the internal vocabulary and cultural norms of the organization.
- Step 5: The team will ensure that all stakeholders are made aware of the risk appetite associated with the procurement at hand by clearly communicating the risk appetite statement at routine intervals throughout the procurement lifecycle.

5.2.1 Key Risk Indicators

Key risk indicators are informative gauges in the risk management process.[28] Organizations will identify many risk indicators for each project. However, key risk indicators are those that carry significant importance and weight with respect to alerting the organization to the most risky and vulnerable areas of the project.[29] Beyond the traditional key risks of any IT system (e.g., security access, user acceptance, business continuity of operations, etc.), the key risk indicators that are

[27] The team responsible for developing the organizational risk appetite for each procurement should consist of individuals in the organization with subject matter expertise, responsibility for establishing risk thresholds, and accountability for risk management outcomes.

[28] Special Competitive Studies Project & Johns Hopkins University. (2023). *Framework for identifying highly consequential AI use cases*. https://www.scsp.ai/wp-content/uploads/2023/11/SCSP_JHU-HCAI-Framework-Nov-6.pdf

[29] https://www.auditboard.com/blog/how-to-develop-key-risk-indicators-kris-to-fortify-business/

unique to AI systems can be reduced to two categories, 1) complexity of the AI system and 2) scale of the potential impact on the population. These are discussed in detail below.

Complexity of the AI System

Because AI systems have the ability to exceed a human's capacity to apply precise computational understanding of the system's output, the human's ability to fully and accurately validate the output can lead to over-trust in that output.[30,31] If the output is faulty, the risk in using the system goes up. Further, no two AI systems are the same. The reliability or likelihood of output failure of one AI system can yield a risk profile that may only require simple light-touch risk controls while another AI system could require a high degree of aggressive/progressive risk controls.

The "complexity of the AI system" key risk indicator serves as a methodological proxy for estimating the likelihood of output failure (e.g., output that is inaccurate, biased, unfair, unequal, etc..). By using systematic and quantifiable means, the RMF PAIS 1.0 improves upon subjectivity of guessing the likelihood of the risk occurrence as is common in other risk frameworks. For example, systems that use highly explainable[32] models with high-quality, well-suited data and relatively few features in the model are considered low complexity AI systems and are therefore, easier to interpret, it can be estimated that this system would be less risky in certain use cases.[33] On the other hand, systems that involve a multiplicity of data points originating from open or uncontrolled datasets, neural network models that operate as opaque, black-box algorithms with many sets of features, weights, and biases, and/or use of data from complex domains (e.g., medical diagnosis data) are estimated to be highly complex AI systems because they are difficult to understand, explain, and rely upon for interpretable output making error acceptance more probably and therefore riskier in certain use cases.[34] In other words, these systems contain more "unknown unknowns," which greatly elevate the level of risky outcomes that require more aggressive risk treatments and controls.

Scale of the Potential Impact on the Population

The second, and equally important key risk indicator in AI systems, is the potential for the decision output to impact individuals in negative, unfair, unequal, biased, and/or otherwise harmful way that may infringe upon their safety and/or fundamental human and civil rights.[35,36] Many factors can lead to negative outcomes when AI systems support and/or make decisions that impact humans. One of the most significant risks in high-risk domains is that these inequities can have

[30] See Footnote 4.
[31] Tartaro, A., Panai, E. & Cocchiaro, M.Z. (2024). *AI risk assessment using ethical dimensions. AI Ethics*. https://doi.org/10.1007/s43681-023-00401-6
[32] Miller, C. L. & Waters, G. (2023). AI procurement: Explainability best practices. The Center for Inclusive Change. https://www.inclusivechange.org/ai-governance-solutions/ai-explainability
[33] See Footnote 21.
[34] See Footnote 11.
[35] Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. https://doi.org/10.6028/NIST.SP.1270
[36] Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2023). *Personalization within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalized feedback*. https://doi.org/10.48550/arXiv.2303.05453

life altering and detrimental impacts on a person's life.[37] Consequently, the more people the AI system could negatively impact (such as in the case of welfare benefits or Medicaid denials), the more resources and expertise are required to identify the source of the issues, adjudicate the issues, and rectify (or reverse) the issues in a timely manner —all of which elevates the risk level for the organization. The fewer administrative users involved in effecting the system's outcomes, the fewer resources that are needed to identify, adjudicate, and rectify the issues, which by default means that the risks are more manageable from an administrative and operational (process) perspective.

Further explanations and examples for the two key risk indicators are provided throughout the risk score card section below.

5.2.2 Risk Appetite Scorecard[38]

Establishing a risk appetite is a slightly subjective exercise. As noted by PriceWaterhouse Coopers, "Some elements [of risk appetite] can be quantified but ultimately it is a question of judgement."[39]

Hence, the use of the risk scorecard is designed to apply a quantitative reasoning approach in order to guide essential critical thinking needs around the two key risk indicators that apply to AI systems.

The score card contains two parts. Part 1 includes an assessment of harms that may impact the population. Part 2 contains an assessment of complexities that the team is willing to accept (or not accept) with the AI system. The score card exercise is meant to be completed by each team member individually. The individual results should be compared, discussed, and then the team should agree on a final score card that reflects a consensus-based risk assessment score card.

The final risk assessment score card will consist of a score for the Population Impact and a separate score for the AI System Complexity. These scores will then be applied to the Risk Matrix, which is further discussed in Section 5.3 below.

5.2.2.1 Population Impact

Part 1 of the score card addresses the potential impacts that the notional AI system is likely to have on the target population. For clarity, the target population means the population for which the system is designed to help or serve. For example, in the case of a welfare eligibility system, the target population would be individuals who may need welfare benefits (not the adjudicating agency). In the case of a system designed to determine the amount of disposable medical supplies shipped to a hospital for just-in-time stock maintenance, the target population would be hospital staff and patients.

The potential harms that are considered on the score card include harms to an individual's health and safety, emotional/psychological, loss of opportunity, economic impact, loss of liberty, and loss

[37] See Footnote 16.

[38] A redacted sample of the score card is available in Appendix B. A full version can be obtained through the AI Procurement Lab upon request.

[39] Barfield, R. (2020). *Risk appetite – How hungry are you*? PriceWaterhouseCoopers. https://www.pwc.com/gx/en/banking-capital-markets/pdf/risk_appetite.pdf

of privacy.[40] In some cases, an AI system may pose no harm at all. In other cases, an AI system may pose multiple types of harms. Scores on the score card should not be limited, but rather should be used to indicate all relevant harms.

Each harm is evaluated and given a score against four dimensions—severity of the harm, population scope, vulnerable populations, and direct/indirect impact.

- **Severity of harm**: The team should agree on the definitions of each severity level prior to the scoring exercise. For example, the definition of a "minor" harm to one team member may have a different interpretation to another. As an example, one team member may ascribe a "life-altering harm" for a scenario where a disabled individual is algorithmically denied 50% of his disability compensation each month that he had been receiving for the past 20 years, but another team member may only consider that change to be a "major" harm. This is where prior understanding of the legitimate business problem becomes highly relevant. The historic data and stakeholder feedback from the root cause analysis of the business problem can provide helpful information throughout this process.
- **Population scope**: This factor is intended to carefully consider the size and scope of the target population relative to the size and scope of the greater population. For example, in an education setting, if the AI system is to monitor students' internet activities, then the target population is the entire student population. As opposed to an AI system design to help 8th grade students prepare for an upcoming standardized test, which would only include a narrow scope of students within the K-12 student population.
- **Disproportionality**: The focus of this evaluation criteria is on diverse, marginalized, and vulnerable populations.[41] Of the population scope (defined in bullet 2 above), what percentage of that population includes underrepresented and vulnerable individuals?[42] What we know about systemic biases is that they are exceedingly difficult to root out of automated systems. Hence, we need to raise the level of awareness in risk potential when underrepresented and vulnerable populations are providing inputs to an AI system. The challenge is to proactively identify training datasets that do not robustly represent the actual users of the system and to do so well **in advance** of deployment.
- **Indirect or Direct Impact**: AI systems can support or make decisions that directly or indirectly impact people. For example, distributing just-in-time supplies to a hospital may have an indirect impact on an individual if there is an existing backup plan to borrow overflow stock from a nearby facility in the event of a temporary supply chain disruption. On the other hand, a decision to deny welfare benefits has a very direct impact on an individual. When the impact is direct, the risks are amplified. Hence the risk score card

[40] The RMF PAIS 1.0 does not address environmental risks that have an impact on the depletion of natural resources over time, but rather it focuses on risks that have immediate and discriminatory impacts on individuals' personal safety, security, livelihood, wellness, and wellbeing.

[41] Diverse, marginalized, and vulnerable populations include but are not limited to individuals or groups that may be statistically and/or historically disadvantaged and/or protected by laws and regulations (e.g., race, religion, gender, sexuality, ethnicity, family status, military status, medical status, disability, socio-economic status, etc.).

[42] Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges, and vulnerabilities. *Journal of Responsible Technology, 4*(100005). https://www.sciencedirect.com/science/article/pii/S2666659620300056

calls for all scores on that row to be doubled (multiplied by two) when calculating the risk score for that particular harm.

In summary, the scoring of each parameter for Part 1 should be summed at the bottom of the scoring worksheet in order to arrive at a total that can be plotted on the Population Harms axis (X axis) in the Risk Appetite Matrix (described in Section 5.3 below). As a reminder, it is important for each individual of the team to complete an individual score card first. The team should then compare their individual results and negotiate a consensus-based scorecard that is used to determine the organization's risk appetite for the procurement at hand.

5.2.2.2 AI System Complexity

Part 2 of the score card addresses the complexities that the team is expecting to encounter within the AI system. Given that a vendor will not have a perfect solution, and the team will have to compromise on the chosen AI system, this part of the score card takes on a different type of thinking. Here the team will consider the risks that they are willing or not willing to accept and/or known probable risks that may be unavoidable and must be controlled through strategic mitigation efforts. In this part of the score card, the team will consider aspects of the data that goes into the decision; how complicated the decision is, even for a human to make; what they want and expect out of the algorithmic models; how explainable the system should be; and compliance with jurisdictional laws and regulations. All of these parameters combined determine the overall complexity of an AI system.

- **Data Origin**: Where the data comes from can influence the quality of the data, which impacts the complexity and subsequent risks within the system.[43] This parameter reviews considerations between internal and external data origins. It also considers how much jurisdiction the developer has over the data and the ability to establish and maintain consistent data quality and data security.[44]
- **Data Validity**: Excellent data validity[45] enables more reliable and constant output. Poor validity in the data can lead to unreliable, biased, and unfair output. This parameter assesses data validity for the procurement at hand including the need for the data to be fit for purpose,

For example, if the organization has a robust, high-quality, fit-for-purpose, and well-tested dataset, the organization may insist on *only* using its own data in the new AI system and prohibiting the vendor from using any other external data. This would create a low-risk scenario and is an admirable goal.

However, the team may encounter vendors that have already trained their systems on external data and cannot abide by such a condition. In that case, the team may have to compromise and establish a stringent risk control to address the issue.

The score card helps the team identify their risk appetite, recognize when an aspect of the proposed solution will exceed their risk tolerance, and immediately identify that a control mechanism must be established in order to realign the solution with the risk appetite.

[43] Yang, X, Wang, X, Zhang, Q, Petzold, L, Wang, W. Y., Zhao, X, & Lin, D. (2023). *Shadow alignment: The ease of subverting safely aligned language models*. https://arxiv.org/pdf/2310.02949.pdf

[44] Data privacy injuries can occur in systems using sensitive personal information even if the system is secure and producing accurate outputs. AI systems are likely to violate growing privacy statutes if they do not keep pace with evolving notice and consent, data minimization, and other legally prescribed data requirements.

[45] EU AI Act. (as of December 2023). See Article 10. Data and Data Governance. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206

consented, representative, robust, accurate, and complete.

- **Input/Output Relationship Complexity**: The complexity of the relationships between inputs and outputs[46] runs on a broad continuum from observable inputs that have simple linear relationships to corresponding outputs to highly complex, sub dynamic, incursive/recursive, pattern-driven input/output relationships. Complexity theory[47], Luhmann's theory of social systems[48], and the simulation of Robert Rosen's anticipatory system[49] inform our approach in defining the complexity parameters among input/output relationships. As such, the parameters assessed in this section are designed to identify the level of complexity within the expected requirements of the procurement at hand.
    - **Expertise**: The parameter contemplates the level of technical and knowledge-based complexity that is resident within the decision. For example, could the decision easily be made by an untrained individual or does the decision require a team of expertly trained Ph.D. scholars that understand highly technical or scientific concepts.
    - **Causality**: The causality parameter evaluates the relative simplicity or complexity of the inputs that lead to the output(s). The more inter-related causal patterns in the inputs that are required to derive an output, the more complex the AI system.
    - **Linearity**: In some cases, outputs can have proportional relationships to inputs in a direct and linear manner. In other cases, the features of a system can cause minor, moderate, and/or major changes in the weights of certain inputs causing changes to outputs in widely disproportionate ways. This parameter analyzes the expected needs for proportionate linearity within the decision as a matter of risk management.
    - **Reducibility**: Some decisions are easily understood because data is minimized and patterns are easily recognized (e.g., all forms of chairs are clearly labeled, and the user needs a wheelchair that the system correctly produces in the output). The data inputs and output in this scenario are easily reducible. When working with neural networks that use machine-labeled data across 70 billion parameters, data labeling and reducibility are highly complex processes. This evaluation parameter considers the expected system type and the reducibility therein.
    - **Solvability**: Problem solvability also runs along a continuum. On one end of the spectrum, predictable recurrent inputs are processed in a systematic way and produce predictable recurrent outputs that contain little to no variations. On the other end of the spectrum, solvability may require dynamic interactions with

[46] Poli, R. (2019). A note on the difference between complicated and complex social systems. *Cadmus Journal, 2*(1). https://www.cadmusjournal.org/files/pdfreprints/vol2issue1/reprint-cj-v2-i1-complex-vs-complicated-systems-rpoli.pdf

[47] Turner, J. R., & Baker, R. M. (2019). Complexity theory: An overview with potential applications for the social sciences. Systems, 7(1), https://www.mdpi.com/2079-8954/7/1/4/htm?trk=public_post_comment-text

[48] Niklas Luhmann (1982) The world society as a social system. *International Journal of General Systems, 8*(3), 131-138. https://doi.org/10.1080/03081078208547442

[49] Leydesdorff, L. (2005). Anticipatory systems and the processing of meaning: A simulation study inspired by Luhmann's theory of social systems. *Journal of Artificial Societies and Social Simulation, 8*(2). https://jasss.soc.surrey.ac.uk/8/2/7.html

disparate patterns of inputs that make each output unique, such as in complex medical diagnoses. This evaluation parameter considers the expected solvability of the business problem that the procurement solution is seeking to address.

- **Model Interpretability**: This assessment parameter considers the expected interpretability of the system to be procured. The process, technical, and ethical choices of a system, including datasets, algorithms, features, weights, etc., can determine how well the output from the decision can be reverse engineered and "interpreted" by a human. When complex algorithms and/or systems constructs (i.e., supply chain complexity) are chosen, interpretability loss occurs, which thereby makes it difficult to understand how a decision was made and which harms may have been escalated in the process.[50, 51]
- **System Performance**: In certain circumstances, it is critical that a system performance achieves near 100% accuracy (e.g., medical diagnosis, chemical analysis of the public water supply, etc.). In other cases, a lower threshold may be acceptable (e.g., counting traffic at an intersection to determine if the timing of a traffic light needs to be adjusted). In addition, AI presents many novel security risks.[52, 53] In all cases, we expect systems to achieve high levels of fortitude and resilience in its security posture. This parameter assesses the expected performance of the system to be procured in terms of desired accuracy.
- **System Explainability**: Various stakeholders require various forms of system explainability. For example, end users require explanations to be written in plain language and pulsed to them as they travers the system so they can absorb the explanations at a reasonable pace.[54] AI auditors require highly technical system documentation during prescribed audit engagements under highly restrictive non-disclosure requirements. This assessment parameter considers the expected level of explainability of the system to be procured.
- **Legal and Regulatory Complexity**: As the legal and regulatory landscape evolves for AI systems, system complexity and the inherent risks will also evolve. This assessment parameter considers the expected consideration and incorporation of jurisdictionally

[50] Widder, D. G. & Nafus, D. (2023). Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data and Society, 10*(1). https://doi.org/10.1177/20539517231177620

[51] Bommasani, R., Kapoor, S., Klyman, K., Longpre, S., Ramaswami, A., Zhang, D., Schaake, M., Ho, D. E., Narayanan, A., Liang, P. (2023). *Considerations for governing open foundation models*. https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf

[52] Vassilev, A., Oprea, A., Fordyce, A., Anderson, H. (2024). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2023

[53] Department for Science, Innovation, and Technology. (2023). *Capabilities and risks from frontier AI: A discussion paper on the need for further research into AI risk*. https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf

[54] United Kingdom, Information Commissioner's Office & Alan Turing Institute. (2023). *Explaining decisions made with AI*. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/

applicable laws and regulations that must be complied with in the system to be procured. (e.g., U.S. federal, state, local laws).[55]

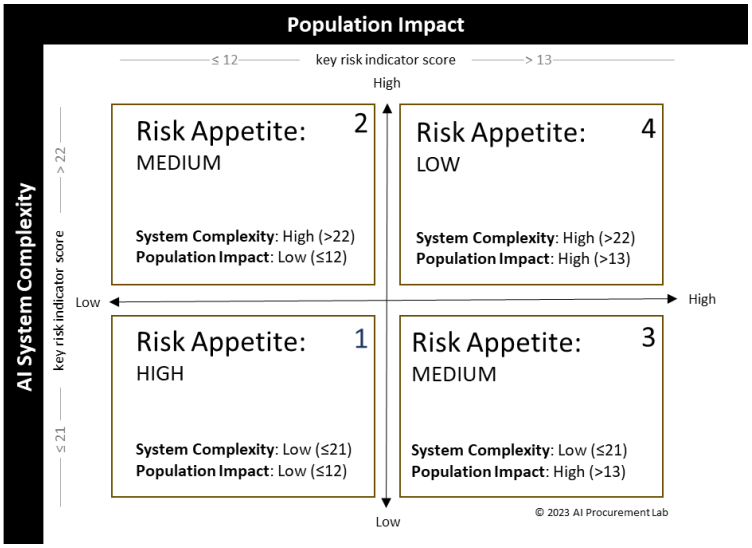
The scoring of each parameter should be summed at the bottom of the scoring worksheet in order to arrive at a total that can be plotted on the AI System Complexity axis (Y axis) in the Risk Appetite Matrix (described in Section 5.3 below). As a reminder, it is important for each individual of the procurement team to complete an individual score card first. The team should then compare their individual results and negotiate a consensus-based scorecard that is used to determine the organization's risk appetite for the procurement at hand.

5.3 Risk Appetite Matrix

A risk appetite matrix is used to enable the Procurement Team to identify a simplified *Risk Appetite Statement*, which is the guiding reference statement that is referred to through the balance of the procurement lifecycle in order to maintain the desired risk tolerance for the organization. The 2x2 Risk Appetite Matrix contains four quadrants and includes an X and Y axis representing the two key risk indicators present in high-risk AI systems. (See Figure 3 below.) The Y axis is based on the complexity of the AI system, and the X axis is based on the scale of the impact on the population.[56]

Figure 3

Risk Appetite Matrix



[55] Holistic.ai. (2022). *U.S regulation of AI and algorithms: Federal, state-level, and local approaches*. https://www.holisticai.com/blog/whitepaper-us-ai-regulation

[56] European Union. (2023) Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts, Section 2. legal basis, subsidiarity and proportionality, Paragraph 2.3 proportionality. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206

In order to determine which of the four quadrants most appropriately represents the risk appetite for each specific procurement, we will use the scores derived from the key risk indicator score card as noted in Sections 5.2.2 above.

5.3.1 Interpreting the Quadrants

The purpose of a risk appetite is to establish an understanding of how much risk an organization is willing to accept for a given situation.[57] The way in which we ensure that this level of risk is not exceeded is to establish risk controls.[58] Simply put, the Risk Matrix is designed to guide procurement teams on which areas of the procurement may require greater scrutiny and therefore, greater risk control prioritization. Examples of procurement types that may align with each quadrant can be found in Appendix C.

Quadrant 1: HIGH Risk Appetite (*Population Impact ≤12, System complexity score ≤21*)

Systems that fall in this quadrant are projected to impact only a small percentage of the population and have low complexity (e.g., they use models that are easily explained, the data is minimized, the data sources are known, safe, robust, representative, etc.). As a result, the organization is likely in a position to tolerate more risk in the procured system.

- Risk controls in this scenario will focus on ensuring that the parameters of the system adhere to the understanding that the system is simple and safe – always.
- Contract monitoring tactics will place extra scrutiny on version control and incident management tracking to confirm that the system parameters are not transforming into a more complex system over time that introduces greater risks and harms.

Quadrant 2: MEDIUM Risk Appetite (*Population Impact ≤12, System complexity score >22*)

Systems that fall in this quadrant are projected to impact only a small percentage of the population and have low complexity (e.g., they use models that are easily explained, the data is minimized, the data sources are known, safe, robust, representative, etc.). As a result, the organization is likely in a position to tolerate more risk in the procured system.

- Risk controls in this scenario will focus on ensuring that the parameters of the system adhere to the understanding that the system is simple and safe – always.
- Contract monitoring tactics will place extra scrutiny on version control and incident management tracking to confirm that the system parameters are not transforming into a more complex system over time that introduces greater risks and harms.

Quadrant 3: MEDIUM Risk Appetite (*Population Impact >13, System complexity score ≤21*)

Systems that fall in this quadrant are expected to have a lower level of system complexity but are likely to impact a much greater portion of the population.

[57] See Footnote 6.
[58] See Footnote 17.

- Risk controls in this scenario will focus on identifying and prioritizing risks related to the ethical choices embedded within the system. Risk controls will also ensure that everyone involved understands how the system manages fair and equal outcomes.
- Rigorous incident management and algorithmic drift monitoring will be essential elements of contract monitoring to ensure the system stays within the prescribed risk tolerance.
- Further, any changes in the vendor's leadership and/or AI governance practices could change the risk profile of this type of use case. Hence, it will be important to maintain routine reporting, attestations, and audits to confirm that those key elements remain sound throughout the life of the contract.

Quadrant 4: LOW Risk Appetite (*Population Impact >13, System complexity score >22*)

Systems that fall in this quadrant are expected to not only impact a large portion of the population but will do so using a highly complex system.

- The stakes in these use cases are high, and thus, the risk control requirements must match the seriousness of the situation. All ethical and technical choices must be scrutinized; acceptable performance levels must be well defined; and close and continuous monitoring of all KPI's and adverse incidents throughout the system life cycle will be essential safety requirements.
- These types of AI systems will also require a "circuit breaker" mechanism whereby the system can be shut down in a matter of minutes if the performance exceeds the risk tolerance level.
- It is important to note that in certain use cases, deviations from these standards and practices could create life-altering and irreversibly detrimental conditions for individuals.

Once you have plotted your risk appetite scores from your consensus-based scorecard onto the risk appetite matrix, you can move on to adopt a risk appetite statement and begin using the recommended levels of control as you assess the proposed systems for the procurement at hand.

5.4 Risk Appetite Statements

The purpose of a risk appetite statement is to serve as a guide for any individual responsible for developing the system requirements, evaluating vendor proposals, negotiating the vendor agreement(s), and monitoring the contract(s).

The importance of a risk appetite statement is to provide the team with a clear and **shared understanding** of the risk level that is acceptable for the procurement at hand in order to help the team more readily identify risks that fall outside of the desired risk tolerance. Unacceptable risks will need to be controlled or mitigated in order to maintain alignment with the risk appetite.

Appendix D provides four risk appetite statements that align with the risk matrix noted in Section 5.3 above. A risk appetite statement may be adopted as is or adjusted to align more closely with organizational culture, language, and other needs. Risk appetite statements should be broadly published to all members of the procurement team and reiterated in a routine cadence throughout the entire procurement lifecycle.

6.0 STEP 2: Risk-Aware Solicitation Requirements

Step two of the RMF PAIS 1.0 calls for the development of responsible requirement that define the sought-after solution. The organization should develop requirements for the AI system in a manner that aligns with the corresponding risk appetite. In other words, the organization should take care not to request a system that would invite risks greater than what the organization is willing to accept in the context of the risk appetite. Engaging all stakeholders[59] in the development of responsible requirements and conducting an AI impact assessment[60] can help identify and mitigate harms and risks prior to finalizing and publishing a request for bids on the desired AI system. Said another way, organizations should do everything possible to avoid setting the vendors up for failure and/or encouraging risk embeddings within the AI system that can otherwise be avoided by conducting their own critical assessment in advance.

7.0 STEP 3: Risk Assessment

Once the risk appetite is set for the procurement and responsible requirements have been established, we can move to step three in the framework, which involves a rigorous risk assessment of the vendors and their proposed AI systems based on the expectations set by the risk appetite.

Every organization chooses methods to manage risks. Vendors are no different. Hence, vendors may rely on a panoply of existing risk frameworks to ensure risks are effectively controlled at the enterprise level and the solution level. Hence, when the team is evaluating and assessing vendor governance and product development practices in order to identify potential risk exposures, it is important to look for evidence of vendors' using familiar risk management frameworks.[61] (See Appendix E examples of relevant risk management frameworks applicable to AI system developers.) The analysis conducted during step three of the RMF PAIS 1.0 will help procurement teams identify gaps in vendors' organizational AI governance practices and identify risk exposures within the vendors' proposed AI system—particularly whether or not those practices and systems meet or exceed the procuring organization's pre-determined risk appetite.

7.1 Risk Assessment and Mitigation Mapping

The ultimate deliverables resulting from the risk assessment process are risk treatments and controls. Vendors face strong incentives to highlight system benefits and downplay their system risks, which means they may not provide a fully accurate or holistic account of their system's risk profile. While AI benefits are important and should be appropriately captured for evaluative effect, as a buyer, it is equally important to ensure that risks are controlled to the maximum extent in an effort to avoid exceeding the pre-determined risk appetite. Hence, during the risk assessment process (aka – solicitation response review phase), organizations should diligently document two

[59] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[60] Stahl, B. C., Antoniou, J., Bhalla, N., et al. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review, 53*(2023), 12799–12831. https://doi.org/10.1007/s10462-023-10420-8

[61] Narayanan, M. & Schoeberl, C. (2023). *A matrix for selecting responsible AI frameworks*. Center for Security and Emerging Technology. https://doi.org/10.51593/20220029

aspects of identified risks using a risk register (depicted in Appendix F).[62] These two aspects include:

1. *Risk Identification for the Purpose of Risk Acceptance/Sharing/Reduction*: Organizations should use the assessment process to methodically identify risks within vendor governance practices and/or AI system-specific gaps and concurrently map the risks to mitigation strategies. (Note: Final mitigation tactics will be negotiated during the contract negotiation phase.) An example of risks and relevant mitigation tactics can be found in Appendix F. Mitigation strategies can include:
    a. Accept. "Do nothing" or accepting the known risk.
    b. Mitigate. Reducing the risk through a prescribed action plan designed to lower the threat level through automated or manual means, curtail the frequency of occurrence, increase monitoring efforts to capture the issue in rapid response modes, etc.
    c. Share. Establish coordination of efforts with the vendor(s) to mitigate the risks.
2. *Risk Identification for the Purpose of Risk Elimination*: Organizations may determine that some vendors or systems represent extreme risks that cannot be sufficiently cured. In these cases, the organization may choose to use the risk appetite as a barometer to decide if eliminating the vendors/solutions that exceed the allowable acceptable risk level for the procurement is a prudent course of action.

8.0 STEP 4: Risk Controls

Next to the risk appetite, meaningful risk controls are at the core of every risk management framework. These controls fall into two categories. Some controls are fixed and constant, as they apply to organization overall. Other controls take on dynamic characteristics in order to conform to the unique properties within each procurement.

8.1 Organizational Governance Risk Controls

Every organization must have a foundational set of risk controls as part of their organizational readiness to manage AI systems. These controls commonly consist of AI literate employees; responsible AI policies, practices, and procedures; an organizational understanding and respect for legal and regulatory requirements; and accountability structures with a commitment to communicating risks when they arise.[63] The absence of an organizational structure and culture primed for success will result in ineffective and unsuccessful risk management. For example, in effective risk management of high-risk AI systems, one of the mitigations that the organization manages is the speed with which the organization addresses and/or reverses off-based, inappropriate, false, inaccurate, unequal, or unfair AI outcomes. If the organization is lacking in its AI governance on these types of organizational policy and human-dignity sensitivities, harms will go unmitigated, and the deployment of the chosen AI system will yield unaddressed residual

[62] Leva, M. C., & Sheehan, R. (2019). *Developing a risk register to deliver risk intelligence*. Chapter 6, p. 105-125. Routledge, London, UK.
[63] https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook/Govern

risks. Hence, organizational AI readiness is an essential cornerstone to responsible AI risk management control.

8.2 AI Procurement Risk Controls

Risk controls that are specific to the AI system being procured must be inserted into the contract(s) with the AI system vendor(s) in order to ensure both risk management and risk accountability are clear for all parties. Here, there are three lines of defense (LOD) in established contract clauses related to risk controls. In the first LOD, the procurement professional should lean on standard clauses that are widely applicable to all types of AI systems. The second LOD is related to the unique risks presented by the chosen vendor/solution. Finally, the third LOD addresses issue that may occur if a mitigation is breached once a system is deployed. Further explanations are provided for each LOD below.

8.2.1 Standard AI Clauses.

Several themes have begun to emerge with respect to AI system contract terms designed to address risks that are common and prevalent across most AI systems regardless of system domain or use case. The City of Amsterdam has led the way in developing a full scope of recommendations that cover many pressing risk mitigation and management control clauses related to AI systems.[64,65] These clauses include topics such as:

- Purpose: Scope, intended uses, known misuses, etc.
- Data: Avoidance of social constructs in data, data rights, allowable uses, data retention
- System: Performance quality, value chain management, transparency, interpretability, explainability
- Monitoring: Performance, incident management SLA's, KPI's, upgrade approvals/lock-in management, periodic audits

"AI Procurement: Essential Considerations in Contracting" provides a synthesis of the City of Amsterdam's terms and conditions.[66] In addition, since being published in 2021, the European Commission began working with the City of Amsterdam's work to update and develop clauses that align with the EU AI Act.[67]

8.2.2 Vendor Specific Risk Mitigation Clauses

Given that each procurement is designed around a unique use case, specific mitigations and controls should be negotiated and embedded into each contract that are relevant to the risks uncovered for the chosen vendor(s)/solution(s) during the risk assessment process. Customarily, risk controls include either accepting the risk as-is, reducing the risk through various mitigation strategies, or avoiding/eliminating the risk.[68] As such, this step of the RMF PAIS 1.0 requires the

[64] https://www.amsterdam.nl/innovation/digitalisation-technology/algorithms-ai/contractual-terms-for-algorithms/

[65] Miller, C. L. & Waters, G. (2023). *AI procurement: Essential considerations in contracting*. The Center for Inclusive Change. https://www.inclusivechange.org/ai-governance-solutions

[66] https://www.inclusivechange.org/ai-governance-solutions/ai-contract-clauses

[67] https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai

[68] https://csrc.nist.gov/glossary/term/risk_response

procurement team to use the risk register to finalize the risk treatments that are most desirable for each identified risk for the chosen vendor(s)/solution(s).

When developing contract terms in the context of actionable risk monitoring and management, one of the most important aspects is to make sure that the contract language is meaningful, specific, measurable, achievable, relevant, and time bound.[69] One of the principles reasons for this approach is to ensure that the terms can be translated into metrics and key performance indicators (KPIs) that are readily trackable during the life of the contract in a way that ensure the AI system is performing within the organization's prescribed risk tolerance.

Once the clause language has been developed by the procuring organization to address all of the risk mitigation tactics (including relevant measurable and timebound goals), further conversations will be necessary with the vendor(s) in order to negotiate, modify, and adjust the mitigation tactics and metrics to ensure that they are equally fair and reasonable for the vendors(s). The final negotiated decisions should be codified into enforceable contract terms.

8.2.3 Risk Tolerance[70] Breach Terms

Metrics and KPIs identified in Section 8.2.2 are good for keeping an eye on the system, but if the system exceeds the risk tolerance, additional risk measures will need to be clearly defined and understood by the parties. Hence, risk tolerance breach terms should be included in all AI system contracts to address short-term triage (e.g., circuit breaker triggers) and long-term cure action steps that are contractually required if a metric is breached. While it may not be possible or practical to address all risks, it is important to address the risks that may have the most egregious impacts on humans.

9.0 STEP 5: Risk Monitoring

Risk monitoring commonly involves watching metrics to ensure that an AI system is not experiencing concept and/or data "drifting" and producing skewed, undesired, unintended, or unfair outcomes.[71, 72] In addition, post-contract / post-deployment monitoring should also include managing adverse incidents, anticipating and managing any breaches that may occur to predefined tolerance metrics, and controlling system evolutions through the use of approvals and periodic audits.

9.1 Risk Tolerance Metrics

The last step in the RMF PAIS 1.0 focuses on risk monitoring for the purpose of maintaining compliance with the organization's risk tolerance. Ideally, the contract will provide a well-defined set of risk controls that are readily translated into metrics and KPIs for monitoring purposes. Each

[69] Ishak, Z., Fong, S. L., & Shin, S. C. (2019, October). *SMART KPI management system framework*. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (pp. 172-177). IEEE.
[70] Carmichael, M. (2019). Risk appetite vs. risk tolerance: What is the difference? Information Systems Audit and Control Association. https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2022/risk-appetite-vs-risk-tolerance-what-is-the-difference
[71] Webb, G., Hyde, R., Cao, H., Nguyen, H., & Petitjean, F. (2016). Characterizing concept drift. https://doi.org/10.1007/s10618-015-0448-4
[72] https://www.datacamp.com/tutorial/understanding-data-drift-model-drift

metric and KPI should be used to conduct ongoing monitoring throughout the life of the AI system and should be managed to levels at or below its upper limits. The key risk indicators in the risk appetite score card are useful categories to consider when establishing metrics and KPIs for risk tolerance monitoring and management.

9.2 Adverse Incident Monitoring

While metrics and KPIs of an AI system will vary with each use case, all use cases should include adverse incident monitoring. A common KPI for adverse incidents will include the number and type of incidents that occur each day/week/month/year. When applying a risk tolerance measure to a KPI measuring adverse incidents, an organization should consider how many incidents it is willing to tolerate within each incident category and the speed with which the organization is addressing, correcting, and/or reversing any critical incidents that directly impact an individual's civil rights, human rights, and/or dignity.

Fraud System Use Case Example

There are 15 complaints submitted to the adverse incident management system each week describing mistaken identity where fraud notices are sent to individuals because the system mistakenly suspected them of fraud. The individuals were suffering undue harm as their bank accounts were frozen for 7 to 10 days while human reviewers investigated the claims.

How many erroneous notices is the organization willing to tolerate as "acceptable" risk?

What if 15 claims represent 10% of all claims? What if they represent 50% of all claims? What if they represent 100% of all claims?

9.3 Threshold Breach Response

Setting risk tolerance measures should be meaningful, specific, measurable, achievable, relevant, and time bound. Likewise, the same principles hold true when determining corrective actions. Any metric or KPI that is breached should have a corresponding (and contractually agreed upon) trigger point with a clear corrective action plan that seeks to mitigate or further control the escalated risk. From the example above, if 15 claims represent 100% of all claims, the corrective action may trigger an immediate system shutdown with a root cause analysis to occur within 72 hours.

Lastly, accountability in monitoring and management must not be overlooked. Organizations should be clear about assigned responsibilities for monitoring, managing, and addressing the risk tolerance metrics and KPIs to ensure full accountability is achieved.[73]

9.4 System Evolutions and Audits

It is natural for vendors to offer enhancements and upgrades to any IT system after a system is purchased and deployed. However, in the case of socio-technical and high-risk AI systems, any system enhancements or upgrade can alter the inherent risks. As such, every substantial

[73] Financial Stability Board Bank for International Settlements. (2013). *Principles for an effective risk appetite framework*. https://www.fsb.org/wp-content/uploads/c_131011p.pdf

enhancement or upgrade should be critically interrogated in much the same way the original system was evaluated in order to identify new or emerging risks to the organization and/or the end users. Further, since enhancements and upgrades in some systems are not always obvious and model drift (concept and/or data) may happen in unsuspecting areas of the system, it is important to maintain a periodic schedule of AI audits. These audits should be conducted against the AI system to verify and validate that the terms of the contract, particularly the solution requirements, common AI risk terms, and specific risk mitigations terms are being upheld and adhered to in good faith. While no vendor likes to give away their intellectual property, this is easily solved through the use of trained third-party AI auditors operating under strict non-disclosure (NDA) and confidentiality agreements. NDAs are widespread practice between private sector entities all the time to further their own interests in partnerships and combined ventures, there is no reason the same mechanism cannot be used to further the mission of responsible AI procurement.

10.0 Summary

The RMF PAIS 1.0 is meant to provide organizations and procurement teams with an essential tool that classifies the risks embedded within each procurement opportunity for the purposes of risk awareness, assessment, measurement, mitigation, treatment, control, monitoring, and management. The RMF PAIS 1.0 is not meant to supplant any existing procurement practices. Nor is it meant to duplicate any other risk management practices such as ensuring enterprise risks are well controlled, cyber risk is managed, and investment risks are considered with rigorous scrutiny, for example.

While there are many risk management frameworks designed for both broad and narrow applications, many tend to gloss over the importance of establishing a risk appetite at the beginning of the risk management process. This one critical step is necessary to gain consensus and set the guardrails for the entire risk management process. Without a clear risk appetite, assessment, mitigation, controls, monitoring, and management become rudderless and less impactful to the success of the organization’s overall ability to capture, control, and manage tolerable amounts of risk.

In the end, no risk management framework can be successfully implemented without a constant drumbeat of communication to the team involved. Clearly outlining the plan and process in advance is necessary to manage the team’s expectations. Reiterating the risk appetite at each milestone and engaging the team throughout the process with feedback loops will help maximize compliance. Compliance and conformance to the framework must be an ongoing effort to uphold the necessary standards to deliver safe, rights-respecting, and valuable products and services to end users.

APPENDICES

Appendix A: High Risk and Unacceptable Risk Systems

I. High-Risk Systems[74]

Domain*	AI Application Examples
Education	Targeting advertisements, determining access, predicting achievement, evaluate learning outcomes, autonomous test proctoring, AI-driven curriculum delivery, AI-augmented classrooms, AI-recommended learning paths, AI-driven assessments, emotional state detection
Employment	Recruitment, hiring, candidate scoring/ranking, targeted job advertising, skills scraping/assessment, AI-driven interviewing, AI-driven assessments, task allocation, quota setting, automated scheduling, performance monitoring, behavior assessment/monitoring, promotion determination, pay determinations, career path recommendations, succession planning, discipline determination, termination, nudges, emotional state detection
Healthcare	Medication, hospitals, doctors, diagnostics, drug discovery & distribution, family planning, patient care, preventative services, wearables, mental health chatbots
Financial Services	Access to credit, credit scores, background checks, insurance, loans, mortgages, interest, and policy rate fairness/equity
Housing	Background checks, eligibility, affordability, rent controls
Government Benefits	Benefits eligibility (grant, reduce, revoke, or reclaim), e.g., welfare, healthcare, social security, HeadStart, etc.
Public Services	Dispatching of emergency first response services, density/placement/availability of emergency and other public services
Critical Infrastructure	Transportation, communications, emergency services, healthcare, safe food
Essential Utilities	Electric, water, gas, communications
Law Enforcement	Polygraphs, deep fake detection, crime analytics (identifying unknown patterns, hidden relationships, fact interpretation), emotional state detection
Justice and Legal	Recidivism scoring, sentencing determinations, probation risk assessments
Immigration	Risk assessment (security, irregular immigration, health), travel document and supporting document verification, application verification (asylum, visa, residence permits), eligibility checking (asylum, visa, residence permits), emotional state detection
Biometric Identification	Security access points, facial recognition, voice and language processing, speech to text, retina scan, fingerprint scan, DNA swabs, emotional state detection
Safety Components	Autonomous vehicles, autonomous drones, HOV lane monitoring, supply of water/gas/electricity monitoring, AI-driven surgery components

II. Unacceptable Risk Systems / Prohibited AI (EU AI Act, Title II, Article 5)[75]

The EU AI Act has determined that the following types of AI systems pose unacceptable risks and are therefore prohibited systems.

- biometric categorization systems that use sensitive characteristics (e.g., political, religious, philosophical beliefs, sexual orientation, race).
- untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases.

[74] https://www.euaiact.com/annex/3

[75] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

- emotion recognition in the workplace and educational institutions.
- social scoring based on social behavior or personal characteristics.
- AI systems that manipulate human behavior to circumvent their free will.
- AI used to exploit the vulnerabilities of people (due to their age, disability, social or economic situation).

*** Excluded Domains:**

The RMF PAIS 1.0 does not address:

- *Military or weapon systems*. This domain contains inherent and important/legitimate complexities involving national security that can serve as a tradeoff benefit to the risk of destroying human life. As such, this is not a domain that authors are willing to address.
- *Environmental sustainability*. The RMF PAIS 1.0 focuses on socio-technical systems that have direct impacts on humans in terms of their inherent rights (civil rights, human rights, and human dignity). Because the environmental effects of AI systems impact natural resources have an indirect, longer term, and more distributed effect on humans, the authors have chosen not to include this domain in the RMF PAIS 1.0.

Appendix B: Risk Appetite Score Card

Disclaimer: The Risk Appetite Score Card is not meant to be used as legal advice. Usage of the score card should be guided by an AI practitioner possessing a moderate to high level of AI literacy understanding and expertise. See Section 5.2.2.1 for explanations. Intellectual property from this score card has been redacted. Please contact the AI Procurement Lab for further information.

Population Impact Score

	ESTIMATED				
Potential Harm	**Severity** Redacted Content	**Scope** Redacted Content	**Disproportionality** Redacted Content	**Impact (Direct or Indirect)** Redacted Content	**TOTAL=** (Severity + Scope + Disproportionality) x Impact
Health / Physical Injury (Bodily injury, death, exposure to unhealthy agents or physical hazard, medical misdiagnosis, medical access, AI-facilitated violence, AI-error/damage in critical infrastructure systems)					
Emotional/Psychological (Loss of autonomy, invalidation, dehumanization, intrusion on emotional state, emotional/sentiment analysis, IQ analysis, work pace controlled by AI/increased worker stress/job insecurity, distortion of reality, attention hijacking/addition, reputational damage, identity misclassification, identity theft, discrimination, misattribution, AI as a human agent to carry out human-to-AI interactions in critical domains)					
Loss of Opportunity (employment, housing, education, insurance, benefits, utilities, critical services)					
Economic Impact (credit discrimination, price discrimination, devaluation of individual occupation(s), skills atrophy/degradation of human skills over AI skills resulting in lower wages, job simplification resulting in lower wages					
Loss of Liberty (False accusations, misattribution, social control, homogeneity, loss of effective remedy)					
Loss of Privacy (Privacy violation, loss of dignity, loss of anonymity, required participation / forced association, disproportionality of information retained in a permanent record)					
				TOTAL	

Note: Some systems may involve one or more human risks/harms while other systems may not involve any harms to individuals at all. Enter an appropriate score in each box that represents the appropriate score relative to each real or perceived harm that is expected. Enter **0** in the box if the harm is not applicable (aka – Indicating there is no perceived or actual harm expected to impact individual(s) for that item).

AI Complexity Score

See also Section 5.2.2.2 for additional explanations. Please contact the AI Procurement Lab for further information.

Parameter	Scoring Consideration	Estimated Score
Data Origin (publicly available/open, vendor/closed/synthetic, internal/org/operational)	**Select a score based on the guide below (combine scored if applicable)**: INTERNAL DATA - organization's own data **1** = Datasets are **2** = Datasets may have **3** = Datasets have not **4** = Datasets have not/may not EXTERNAL DATA - vendor/closed, public/open-source, or synthetic data [Note: Scoring jumps from 1 to 3 due to increase in risk.] **1** = Datasets are **3** = Vendor uses **4** = Vendor uses **5** = Vendor uses	
Data Validity (Fit for purpose, consented, representativeness, robustness, accuracy, completeness)	**Expertise: Select a score between 1 – 3 based on the guide below**: **1** = All datasets are/will **2** = All datasets are/will **3** = All datasets are/will	
Input/Output Relationship Complexity (expertise, causality, linearity, reducibility, solvability)	**Expertise: Select a score between 1 – 3 based on the guide below**: **1** = Output is **2** = Output is **3** = Output is	
	Causality: Select a score between 1 – 3 based on the guide below: **1** = Cause and effect of the input/output **2** = Cause and effect of the input/output **3** = Cause and effect of the input/output	
	Linearity: Select a score between 1 – 3 based on the guide below: **1** = Every output of the system **2** = The most significant inputs **3** = Outputs are not	
	Reducibility: Select a score between 1 – 3 based on the guide below: **1** = The inputs and how they **2** = The inputs and how they **3** = The inputs and how they	
	Solvability: Select a score between 1 – 3 based on the guide below: **1** = Problems contain **2** = Problems contain **3** = Problems contain	

(continued)

Parameter	Scoring Consideration	Estimated Score
Model Interpretability (Features, parameters, weights, and calculations, and decisions are interpretable)	**Select a score between 1 – 3 based on the guide below:** **1** = the system/models will **2** = the system will **3** = the system will Redacted Content	
System Performance (output accuracy, reliability, performance consistency)	**Select a score between 1 – 3 based on the guide below:** **1** = Output must be **2** = Output must be **3** = Output must be Redacted Content	
System Explainability (technical documentation, administrator training, tickle-through explainability)	**Select a score between 1 – 5 based on the guide below:** **1** = **2** = **3** = **4** = **5** = Redacted Content	
Legal and Regulatory Complexity	**Select a score between 1 – 3 based on the guide below:** **1** = Legal and regulatory **2** = Legal and regulatory **3** = Legal and regulatory Redacted Content	
Total		

Important Note:

According to Vassilev, et. al,[76] when it comes to AI/ML systems, "security and privacy challenges include the potential for adversarial manipulation of training data, adversarial exploitation of model vulnerabilities to adversely affect the performance of ML classification and regression, and even malicious manipulation, modifications or mere interaction with models to exfiltrate sensitive information about people represented in the data or about the model itself." The RMF PAIS 1.0 only focuses on security related to training data given that it is a slightly more mature field of research in terms of risk management.

Risks related to model vulnerabilities are novel and continue to evolve at a rapid pace. The RMF considers model vulnerabilities as assumed risks present in every AI system regardless of use case. Likewise, the RMF considers user interfaces, system architecture and infrastructure, and other ordinary IT system components that are commonly subject to threats and risks as assumed risks present in every AI system regardless of use case.

Hence, model vulnerabilities and ordinary IT system threats/risks are NOT deeply accounted with any significant specificity accounted for in this RMF but should be managed appropriately when conducting risk assessments and risk control activities.

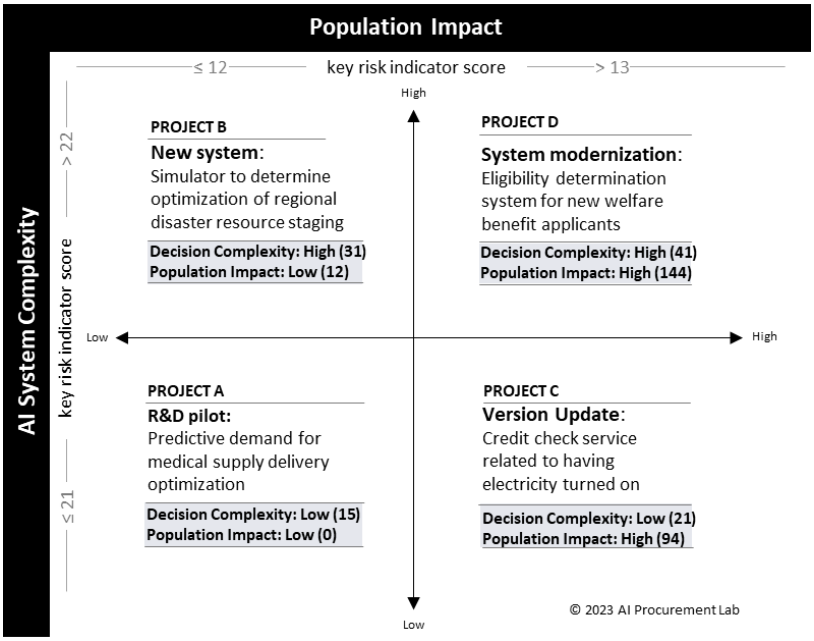
[76] Vassilev, A., Oprea, A., Fordyce, A., & Anderson, H. (2023). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2023

Appendix C: Risk Appetite Estimation Matrix Examples

Identifying a risk appetite is not a perfect science. There is no right answer when determining the risk appetite for each procurement. Prior to conducting a risk appetite assessment, it is critical to gain background information and perspectives from the stakeholders that will be most impacted by the AI system—particularly if vulnerable individuals are involved. This assessment can also benefit from close cooperation with internal stakeholders and guidance from procurement managers that have had experience(s) with AI procurements as well as AI vendors in the target domain of interest (if possible). Again, the primary purpose is to approximate the amount of risk that the organization is willing to accept (or not accept) for each procurement – prior to commencing any evaluation of any vendors or technical system proposals. Figure C1 offers four examples for illustrative purposes:

Figure C1.

Risk Appetite Matrix



Further hypothetical context for each project is provided below:

- **Project A. R&D pilot**: This system is an exploratory research project, which will not impact individuals during the research and development phase. This makes the risk profile low. The data for this system is internally sources, of high-quality, and fit-for-purpose. Further, the algorithms are explainable, and decisions are causal, systematic, and variations are predictable. These characteristics mean that the AI system has a low complexity. With a low AI system complexity and a low impact on the population, the risk appetite for this procurement can be set at high. Obviously, if the organization will need to reevaluate the next procurement when/if the system is moved from an R&D project into a fully deployed

system since the population impact will be substantially impacted by a deployed system, thereby changing the risk profile.

- **Project B. New system**: This system is an advanced simulator designed to simulate the optimization of staging of emergency support assets at regional locations for impending disasters. The purpose of the system is strictly for determining annual budgets and adjusting budgets as needed. The system's impact on individuals is indirect and only has to do with physical assets made available to disaster site workers if funding is fully depleted without an appropriate contingency plan. Hence, the direct impact on the population is relatively low by comparison to other systems. The data for this system is both internally from previous disasters and externally sourced from government and private-managed datasets, some of which are highly technical geospatial, weather, geological, financial, and other specialized data. Further, the algorithms contain complex neural networks, problem solving is non-linear, cause and effect is based on multiple interacting conditions and patterns observed within the datasets, and outputs are not proportional as small changes can lead to significantly different outcomes. These characteristics mean that the AI system has a high complexity. With a high AI system complexity and a low impact on the population, the risk appetite for this procurement can be set at medium.
- **Project C. Version update**: This project is a version update of an existing application system for the electric utility provider. The primary feature of the version includes several modifications to the credit check module. Since this electric utility provider services five major metropolitan areas covering 30 million households and they experience an average of 20% new applications each year, this system will have a high impact on the population. Further, of the new applications that are submitted each year, over half of those applications originate for low-income neighborhoods, which means the vulnerability for this application is increased. The algorithms in the system are explainable and decisions are causal, systematic, and variations are predictable. These characteristics mean that the AI system has a low complexity. With a low AI system complexity and a high impact on the population, the risk appetite for this procurement can be set at medium.
- **Project D. System modernization**: This system will impact approximately 55 million people eligible for and/or receiving public financial compensation welfare benefits. This is a statistically significant portion of the population, and all individuals impacted by the output from this system are considered vulnerable individuals. The AI system will utilize hundreds of complex domain-specific data points to determine initial and ongoing eligibility for compensation and the amount of compensation. The algorithms contain complex neural networks, problem solving is non-linear, cause and effect is based on multiple interacting conditions and patterns observed within the datasets, and outputs are not proportional as small changes can lead to significantly different outcomes. These characteristics mean that the AI system has a high complexity. With a high AI system complexity and a high impact on the population, the risk appetite for this procurement can be set at low.

Appendix D: Risk Appetite Statements

The following statements are notional and should be adapted to match the tone and cultural values of your organization for maximum effect.

Risk Appetite Statement for Quadrant 1: Lower Left, High Risk Appetite

> *We have a HIGH-risk appetite with regard to _____ [insert title of procurement here]. We will tolerate this level of risk for this project only because the project will have a minimal impact on the greater population, there are either no known risks or very limited low-probability risks identified that will impact vulnerable populations. Further, the AI system is expected to be easily interpretable, explainable, understood, reliable, consistent, accurate, and continuously monitored and managed by trained individuals. If these conditions prove to be untrue and the risk tolerance is exceeded, the project must be placed on hold and a reassessment of the true risks must occur.*

Risk Appetite Statement for Quadrant 2: Upper Left, MEDIUM Risk Appetite (Low population impact, high system complexity)

> *We have a MEDIUM risk appetite with regard to _____ [insert title of procurement here]. We will set priorities and implement risk controls in consideration of the complexities within the AI system that pose the greatest probability of risks to the population coupled with a fair and balanced analysis of stakeholder input.*

Risk Appetite Statement for Quadrant 3: Lower Right, MEDIUM Risk Appetite (High population impact, low system complexity)

> *We have a MEDIUM risk appetite with regard to _____ [insert title of procurement here]. We will set priorities and implement risk controls in consideration of the key stakeholder feedback that indicates the greatest risks to the target population coupled with a fair and balanced analysis of the probability of those outcomes.*

Risk Appetite Statement for Quadrant 4: Upper Right, LOW Risk Appetite

> *We have a LOW-risk appetite with regard to _____ [insert title of procurement here]. We have no appetite for harm to the target population. While we recognize that we cannot eliminate all manner of risk and harm, will work to establish, and maintain strong controls to mitigate risks within every aspect of the procurement system.*

Appendix E: Risk Management Frameworks

One or more relevant risk management frameworks being applied by the vendor should be clearly evidenced when assessing vendors and their proposed systems.

Enterprise and Software Risk Management Framework

- Committee of Sponsoring Organizations (COSO) Enterprise Risk Management Framework[77]
    - Note: Most useful as an overall corporate risk governance framework to identify and manage risks across financial investments, physical assets (e.g., buildings, data centers, call centers, etc.), human resources, software systems, etc. for an expansive risk landscape (e.g., capital investment decision making, disaster planning and recovery, system outages, cyber-attacks, individual incident management, etc.)
- NIST 800-30[78] and ISO 27005[79]
    - Note: Most useful for information security and privacy elements that can lead to risks. However, AI has created new attack vectors through the use of public data sets, prompts, and other aspects of machine learning entry points and may require addition levels of risk management.
- IEEE 1012[80]
    - Note: Most useful for verifying and validating system software and hardware quality elements that can lead to risks. However, IEEE 1012 may be more useful in a rules-based system where a system's outcomes are more predictable. In an AI environment where a neural network/black box model can produce inconsistent outcomes and machine learning drift, verification and validation points are a moving target and are more challenging to confirm with certainty. Still, IEEE 1012 is a useful standard for an AI system in terms of documenting and evaluating choices, business rules, and other system and hardware requirements.
- Failure Mode and Effects Analysis (FMEA)
    - Note: Most useful as an **ex-ante** approach to identifying and attempting to address potential failures or issues (and the possible outcomes of those failures) before the event actually occurs.[81]

AI-Specific Risk Management Frameworks and Tools

These frameworks have been developed specifically to identify, assess, treat, and manage risks within AI systems. These frameworks primarily focus on AI safety, security, privacy, fairness,

[77] Nguyen, M & McKeown, P. (2022, January 20). 5 AI auditing frameworks to encourage accountability. Auditboard. https://www.auditboard.com/blog/ai-auditing-frameworks/
[78] https://csrc.nist.gov/pubs/sp/800/30/r1/final
[79] https://www.iso.org/standard/80585.html
[80] https://standards.ieee.org/ieee/1012/5609/
[81] Liu, H.-C., Liu, L., & Liu, N. (2013). Risk evaluation approaches in failure mode and effects analysis: A literature review. *Expert Systems with Applications, 40*(2), 828–838. https://doi.org/10.1016/j.eswa.2012.08.010

equity, bias reduction, explainability, interpretability, transparency, accountability, validity, and reliability.

- NIST AI 100-1[82] (more commonly known as the NIST AI RMF)
- EU AI Act, Title III: Classification of AI Systems as High-Risk, Chapter 2: Requirements for High-Risk Systems, Articles 8 through 15[83]
- ISO/IEC 23894:2023 - Information technology — Artificial intelligence — Guidance on risk management[84]
- ISO/IEC 42001 - Artificial intelligence — Management system [under development as on 11/21/23][85]
- ISO 38507 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations[86]
- US Government Accountability Office (GAO) AI Framework[87]
- IEEE 7000-2021 Standard Model Process for Addressing Ethical Concerns during System Design[88]
- Institute of Internal Auditors (IIA) Artificial Intelligence Auditing Framework[89]
- Singapore Personal Data Protection Commission (PDPC) Model AI Governance Framework[90]
- ForHumanity Risk Management[91]
- IEEE Ethically Aligned Design[92]
- ISO/IEC TR 24027:2021 Information technology Artificial Intelligence (AI) Bias in AI systems and AI aided decision making[93]
- ISO/IEC DTS 12791 Information Technology Artificial Intelligence Treatment of unwanted bias in classification and regression machine learning tasks[94] [under development as on 11/21/23]
- ISO/IEC TR 24368:2022 Information Technology Artificial Intelligence Overview of ethical and societal concerns[95]

[82] United States Department of Commerce (2023, January 26). *Artificial intelligence risk management framework*. National Institute of Science and Technology. https://doi.org/10.6028/NIST.AI.100-1
[83] https://www.euaiact.com/article/8
[84] https://www.iso.org/standard/77304.html
[85] https://www.iso.org/standard/81230.html
[86] https://www.iso.org/standard/56641.html
[87] https://www.gao.gov/products/gao-21-519sp
[88] https://ieeexplore.ieee.org/document/9536679
[89] https://www.theiia.org/globalassets/documents/content/articles/gpi/2017/december/gpi-artificial-intelligence-part-ii.pdf
[90] https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf
[91] https://forhumanity.center/bok/risk-management/
[92] https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
[93] https://www.iso.org/standard/77607.html
[94] https://www.iso.org/standard/84110.html
[95] https://www.iso.org/standard/78507.html

- CISA Supply Chain Bill of Materials Self-Attestation Form (Draft)[96]

Domain-specific risk assessments and frameworks

These frameworks have been developed to identify, assess, treat, and manage risks that are specific to unique domains and use cases. This is only meant to provide a sample of how risk management can and should be applied at the domain level. It is prudent to use a domain specific risk framework in combination with an AI specific risk framework in order to identify and mitigate risks in the most comprehensive way for each AI system use case.

- Accessible Technology Procurement Toolkit[97]
- Department of Energy AI Risk Management Playbook[98]
- National Fair Housing Alliance Purpose, Process, and Monitoring Framework[99]
- Department of Education, Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations[100]
- FDA Medication Risk Evaluation and Mitigation Strategies[101]
- Sex Offender Risk Assessment[102]
- Construction Risk[103]
- Dental Disease Risk[104]
- Fire Risk[105]

[96] https://www.cisa.gov/resources-tools/resources/secure-software-self-attestation-common-form
[97] https://disabilityin.org/procurementtoolkit/section/before-you-buy-investigate-accessibility/
[98] https://www.energy.gov/ai/doe-ai-risk-management-playbook-airmp
[99] https://nationalfairhousing.org/issue/purpose-process-and-monitoring-framework-ppm/
[100] https://tech.ed.gov/ai-future-of-teaching-and-learning/
[101] https://www.fda.gov/drugs/drug-safety-and-availability/risk-evaluation-and-mitigation-strategies-rems
[102] Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33*(2), 287-316.
[103] Taroun, A. (2014). Towards a better modelling and assessment of construction risk: Insights from a literature review. *International Journal of Project Management, 32*(1), 101-115.
[104] Lang, N. P., Suvan, J. E., & Tonetti, M. S. (2015). Risk factor assessment tools for the prevention of periodontitis progression a systematic review. *Journal of Clinical Periodontology, 42*, S59-S70.
[105] Moshashaei, P., & Alizadeh, S. S. (2017). Fire risk assessment: A systematic review of the methodology and functional areas. *Iranian Journal of Health, Safety and Environment, 4*(1), 654-669.

Appendix F: Sample AI Procurement Risk Register

Intellectual property from this score card has been redacted. Please contact the AI Procurement Lab for further information. The following risks are only suggestions. *They are NOT meant to be construed as legal advice*.

This list is simply a representation of many angles by which one can view risks and mitigations. It is not an exhaustive list, nor should this list be applied to a procurement in an exhaustive manner. Some rows may encompass larger risk concepts for the sake of brevity while others may present more narrowly tailored to specific risk concerns that would be appropriate if a vendor demonstrates a specific gap. The NIST AI RMF can be used to identify tactical risk mitigation ideas. Each procurement team will want to create a **risk register** for each vendor that best aligns with the needs of each procurement case. The risk register should consider the scale, scope, and risk appetite for that procurement and the available capabilities of the vendors involved.

Risk Gap		Mitigation Recommendation			
Vendor	**Identified Risk Category**	**Specific Mitigation Tactic**	**Measurement**	**Time**	**Priority (1,2,3,4)[a]**
	Governance roles: Missing AI Ethicist	Redacted content	Redacted content	Submission to buyer within 30 days of contract signature, annually thereafter	
	Diverse leadership	Redacted content	Redacted content	Submission of most current report to buyer prior to contract signature, annually thereafter	
	Diverse staffing	Redacted content	Redacted content	Submission of most current report to buyer prior to contract signature, annually thereafter	
	Policy on Responsible AI	Redacted content	Redacted content	Submission to buyer within 15 days of contract signature, annually thereafter	
	Ethical training	Redacted content	Redacted content	Submission to buyer within 15 days of contract signature, annually thereafter	
	Whistleblower protections	Redacted content	Redacted content	Annual submission to buyer, Buyer to conduct random testing	
	Stakeholder Engagement	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Necessity Assessment	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Proportionality Assessment	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Algorithm Impact Assessment	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Procedural and Ethical Choices	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Version control	Redacted content	Redacted content	Disclosure made to buyer prior to deployment with buyer approval	
	Risk Management Framework	Redacted content	Redacted content	Annual AI Audit Disclosure	

	AI supply chain risk monitoring and auditing	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Testing and acceptance procedures	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Prohibited uses, foreseeable misuse, disuse, and abuses	Redacted content	Redacted content	Submission to buyer within 15 days of contract signature, annually thereafter	
	Adverse Incident Management	Redacted content	Redacted content	KPI/OKR reporting submitted monthly to buyer. Annual AI Audit Disclosure of Random sampling of incidents, inspection of KPI/OKR history	
	Data traceability	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Data fit for purpose, robust, representative, accurate	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Data provenance	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Data generated is strictly used for contract, accrues to, and is owned by the buyer	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Explainability	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Human reviewer training	Redacted content	Redacted content	Submission to buyer within 30 days of contract signature, annually thereafter (highlighting changes based on version updates)	
	Legal and regulatory compliance	Redacted content	Redacted content	Annual AI Audit Disclosure	
	Drift Monitoring	Redacted content	Redacted content	Routine monitoring (daily, weekly, monthly, quarterly, annually)	

a. Each risk identified should be prioritized. 1=highest risk level, 4=lowest risk level. While it is best to attempt to mitigate all risks, it may not be possible as negotiations progress. Assigning a priority level can help guide the negotiations to ensure the highest priority risks are addressed first.

See also:

- ISO/IEC DIS 42001:2022, Annex A: Reference control objectives and controls, https://www.iso.org/standard/81230.html
- Holistic AI's Risk Mitigation Roadmaps: https://holisticai.gitbook.io/roadmaps-for-risk-mitigation/

Disclaimer and Representations:

The authors received no payments, funding, grants, or other sources of income for any part of this document. The authors' opinions are their own. Nothing herein is meant to provide or represent legal advice. Please seek appropriate legal counsel for your procurement projects.

Citation:

Miller, C. L. & Waters, G. (2024). Risk Management Framework for the Procurement of Artificial Intelligence (RMF PAIS 1.0). AI Procurement Lab and Center for Inclusive Change. https://www.inclusivechange.org/ai-governance-solutions/rmf-for-ai-procurement

Feedback:

We would love to see your score cards and hear any feedback you would like to offer that can further the utility of the framework. Please feel free to email any feedback to team@aiprocurementlab.org

AI Procurement LAB

The Center for Inclusive Change