ASIF QAMAR

SUPPORT VECTORS

AI PROCESS AUDIT

Caveat Emptor!

This is currently a work in progress, and incomplete: therefore, please do not circulate it yet.

supportvectors ai lab technical series

www.supportvectors.com

First draft, July 27, 2023.

Typeset in LaTeX.

Contents

1 Executive Summary

Contents

This document is a quick sketch of the AI audit technical process.

1.1 Introduction

This internal use document intends to sketch out the main technical aspects of an AI audit: in particular it answers the following questions:

Why An exploration of why an AI audit must be an essential part of any company venturing into AI development.

What Some specific benefits of an AI audit.

How Sketch of the AI audit engineering process.

1.2 Intended use

This quick write-up should help the Kahoa marketing team in adding more technical context to the collaterals it is creating. It is currently not in form meant to be shared outside Kahoa.

1.2.1 Technical specification of the AI audit procedure

This is not a detailed specification of the AI audit process, but merely its outline. In subsequent weeks, the engineering oriented technical specification of the process will emerge, and be documented separately.

"Happy families are all alike; every unhappy family is unhappy in its own way."

Tolstoy, the opening sentence of Anna Karenina

2 Sketch of the AI Audit

Contents

This chapter contains a sketch of the AI audit process.

2.1 Introduction

AI is the new electricity.

Andrew Ng, Coursera

AI investment in R&D likely to hit $400 billion in 2024.

IDC

AI will contribute $15.7 trillion to the global economy by 2030.

A PwC forecast

AI is arguably the greatest revolution in our lifetimes; it is ushuring in disruptive transformations in almost all industries, and it is hard to imagine an area of human endeavor which will escape AI's impact.

With the emergence of Generative AI and its capabilities, there is an irrational exuberance in the zeitgeist about the potential societal and business benefits that will accrue, with experts forecasting that it will add many trillions of dollars to the global GDP. While the years to come may prove that prognosis correct, such a lopsided perspective misses an essential point, namely, that with power comes attendant responsiblity.

Much like fire, a reckless use of AI can prove catastrophic to an enterprise, and to society at large. Many thoughtful voices are voicing these concerns with urgency.

For an enterprise, however, the core goal has to be the creation of a framework and process that harnesses the power of AI, while at the same time responsibly putting the right guardrails around its use to prevent harm. Laws and regulations have begun to arise – for example, the Europe AI Act, that sets boundaries on what companies can and cannot use AI to create and do. Within the US too, mindful of the potential for harm that AI can cause, various states are passing their own independent laws and regulations in the context of AI. In due course of time, the federal government is likely to follow suit with its own regulations.

Enterprises in almost all industry verticals are grappling with the central question: "What is the impact of AI on our business, and how can we benefit from it?" As a consequence, the current landscape has almost every company engaged with AI in some form. Gartner has described this as an AI maturity model, segmenting companies into five levels of AI adoption maturity.

> Technology adoption will remain a key driver of business transformation in the next five years. Over 85% of organizations surveyed identify increased adoption of new and frontier technologies and broadening digital access as the trends most likely to drive transformation in their organization... Background Companies rank AI and big data 12 places higher in their skills strategies than in their evaluation of core skills, and report that they will invest an estimated 9% of their reskilling efforts in it – a greater proportion than the more highly-ranked creative thinking, indicating that though AI and big data is part of fewer strategies, it tends to be a more important element when it is included.
>
> The Future of Jobs Report 2023

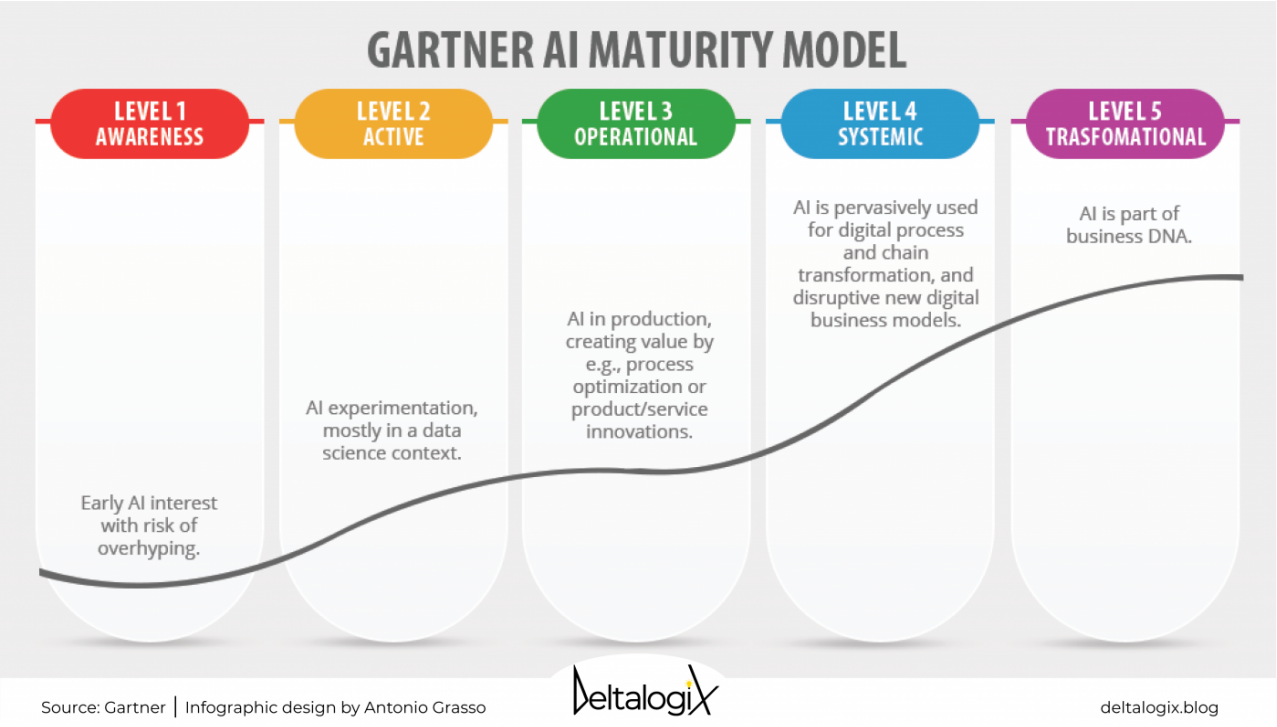


Figure 2.1: The Gartner AI Maturity Model.

2.2 The AI Adoption Dystopia

There is a significant learning curve associated with AI adoption in an enterprise. As companies begin to move forward in the AI maturity curve aforementioned, there is a lot of internal discussions, planning, experimentation and activity, and in some cases a production deployment of an AI-functionality or product. In all of this, a recurring theme is that absence of clearly marked pathways and processes. This has led to most AI projects having less than stellar outcomes, despite the considerable investment and leadership support. The available data paints a rather depressing picture of a dystopia:

Gartner (2019) According to a Gartner survey, approximately 85% of AI projects do not deliver on their intended promises. This can be due to a lack of clear business objectives, absence of necessary skill sets, or unrealistic expectations.

MIT Sloan Management Review (2020) 65% of companies reported that they

have not yet seen value from the AI investments they have made. This could be due to several factors such as the difficulty in integrating AI into existing processes and systems, as well as a shortage of internal AI skills.

IDC (2019) Half of the respondents in their survey reported that they had at least one AI project fail in the previous two years, citing reasons like unrealistic expectations, data quality issues, and a lack of skilled staff.

Capgemini Research Institute (2020) 7 out of 10 organizations reported that they struggled to scale AI implementations. The most common barriers include issues with data quality, data labeling necessary for machine learning, and the lack of a clear data strategy.

What about reports that have emerged in 2021,2022

These statistics suggest that many organizations face difficulties in implementing AI projects, which can lead to a high failure rate.

This leads naturally to the question: how can one form a team of talented AI engineers and data scientists – who presumably undergo years of professional training in computer science – and still have such an alarmingly high failure rate of AI projects?

While the specific causes of failure in each of the projects differ, there are, nonetheless, some common themes that emerge in AI implementation efforts in the enterprise:

Lack of Clear Strategy and Goals Without a clearly defined strategy and measurable objectives, AI deployments can lose direction and fail to deliver meaningful results. AI is a tool to solve business problems, not a silver bullet that works without clear business alignment.

Poor Data Quality AI models rely heavily on data. If the data is of poor quality, incomplete, or biased, it will affect the performance of the AI model, leading to unreliable or inaccurate results.

Insufficient Data Management and Governance AI requires access to diverse and large amounts of data. Without proper data management and governance structures, organizations might face difficulties in maintaining data

privacy, addressing ethical concerns, and ensuring data is used appropriately. This problem is particularly acute in sectors where sensitive data is the norm, rather than the exception. Healthcare is one such place, and Human Resources is another such vertical.

Lack of Skills and Expertise Implementing AI requires a unique blend of skills, including data science, software engineering, and business acumen. If the organization doesn't have the right team with the right skills, the AI project may not succeed.

Scalability Issues Often, AI solutions that work well in a small, controlled environment do not scale up effectively for broader organizational use. For example, a model may have a response time of a half-second for inference, which may be acceptable in the lab. But when the same model is deployed in production, and traffic arises, one may see the significantly higher latency, or the dreaded out-of-memory error on the GPU.

Inadequate Infrastructure A common occurrence is a gross underestimation of the computational needs of AI projects. Unlikely other software applications, AI projects often need math coprocessors such as GPU-accelerators or dedicated Tensor processing units. Once a project has started, an insufficient realization of the needs – an insufficient allocation or availability of compute resources can be a challenge for organizations; for on-premise deployments, there is a latency in acquiring hardware from vendors, and a capital expense consideration. For cloud deployments, there may have been insufficient estimation of the OpEx (recurring operational costs), since compute instances with GPU tend to be expensive.

Insufficient Stakeholder Buy-in For an AI deployment to be successful, it's important to have support from all stakeholders, including top executives, managers, and users. Lack of buy-in can lead to resistance, inadequate resource allocation, and eventual failure of the project.

Lack of Monitoring and Maintenance AI models need to be continuously monitored and updated to ensure they remain effective as new data comes in and conditions change. A lack of robust processes in this context leads to model drift and degradation of model inference performance.

Failure to Consider Ethical Implications AI applications can have significant ethical implications, from privacy concerns to decision-making biases. If these are not adequately addressed, it can lead to a loss of trust and potential legal problems.

Misunderstanding AI Capabilities and Limitations Unrealistic expectations from AI solutions can lead to disappointment and perceived failure. It's crucial to understand that AI isn't a magic solution that can solve all problems without fine-tuning and ongoing management.

2.2.1 Specific risks to an organization

But the alarming failure rate of AI projects is not the biggest risk – there are other factors at play that pose significant risk exposure; these tend to vary by each organization's specific use of AI, the industry in which it operates, and its regulatory environment. However, a significant risk that many organizations face relates to ethical considerations and compliance with legal and regulatory standards.

Ethical Considerations AI systems can pose significant ethical risks if not properly managed. This includes issues related to bias and fairness, transparency, privacy, and security. If an AI system makes decisions that are biased or discriminatory, or if it compromises user privacy, this can lead to significant reputational damage and potential legal consequences.

Legal and Regulatory Compliance Depending on the industry and the region, there may be specific laws and regulations governing the use of AI. For example, in healthcare or finance, there are strict regulations about how data can be used and the types of decisions that can be automated. Non-compliance with these regulations can result in severe penalties.

Data Quality and Management Poor data quality can lead to poor AI model performance, and this risk can be exacerbated by the scale at which AI systems operate. Similarly, improper data management practices can lead to breaches of privacy, loss of critical data, or non-compliance with data protection laws.

Robustness and Security of AI systems AI models are also at risk of being manipulated by adversarial attacks, or they may behave unpredictably in situations that differ from their training conditions.

All these areas need to be thoroughly audited in an AI system to ensure ethical use and minimize risk. However, the relative importance of each area will depend on the specific context in which the AI system is being used.

2.3 Some organizational benefits of an AI audit

Performing an AI audit has several key benefits for a company's AI initiatives:

Transparency and Trust Auditing AI models can ensure that their inner workings and decision-making processes are understandable and transparent. This is crucial for building trust with stakeholders, including customers, employees, and regulators.

Risk Management An AI audit helps in identifying and mitigating risks associated with AI deployment, including ethical, legal, and reputational risks. It can assess vulnerabilities to data breaches, bias, fairness issues, and non-compliance with regulations, enabling the organization to address these proactively.

Improved Performance Auditing can help to identify issues that may be hindering the performance of AI systems. These might be issues with the quality of the data, the selection of features, the type of model used, or the way the model is being trained and validated.

Regulatory Compliance Depending on the industry and jurisdiction, there may be legal requirements to audit AI systems for compliance with specific regulations. Even when not legally mandated, demonstrating a commitment to rigorous auditing can help to maintain good relationships with regulators. For example, in healthcare or finance, there are strict regulations about how data can be used and the types of decisions that can be automated. Non-compliance with these regulations can result in severe penalties.

Ethical Assurance An audit can help ensure that the company's AI systems are being used in a way that aligns with its ethical commitments and values. This includes checking for any bias in the system, ensuring fair decision-making, and respecting user privacy. If an AI system makes decisions that are biased or discriminatory, or if it compromises user privacy, this can lead to significant reputational damage and potential legal consequences.

Operational Efficiency The audit can lead to improved operational efficiency by ensuring that AI models are working as intended and solving the right problems. It can help in identifying over-complicated or resource-intensive processes that could be streamlined.

Stakeholder Confidence Regular audits demonstrate to investors, shareholders, customers, and the public that the company is committed to the responsible use of AI. This can boost their confidence and the overall reputation of the company.

Each of these benefits contributes to the overall goal of ensuring that AI is used responsibly, ethically, and effectively within the organization.

"Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away."

Antoine de Saint-Exupery

3 AI Audit Process Steps

Contents

Here we elaborate upon the details of AI process audit steps.

An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives.

OECD

-

3.1 Introduction

The margin epigraph provides a concise definition of an AI system that we will use extensively. It comes from OECD[1], and its AI policy observatory, which takes an integrated view of the AI ecosystem.

[1] The Organisation for Economic Co-operation and Development. It is an international organization that aims to create better policies that positively impact lives by fostering prosperity, equality, opportunity and well-being for all. (OECD)

Performing an AI audit is a complex process that involves multiple steps to ensure that the AI system is working as expected, is ethical and trustworthy, is reliable and cost-effective, the AI models are performant as SOTA[2] and optimal for the task at hand. Also, we need to ensure that the entire AI system is transparent, explainable and interpretable to the extent possible, bias-free, without adverse impact on any societal section, and is legally compliant. We need to ensure that data privacy is robust, the feature store is well managed, as are security safeguards against malicious attacks or attempts to poison or misguide the models' behavior and predictions.

[2] State of the art. It implies that a particular implementation aligns with the best comtemporary benchmarks, standards and practices.

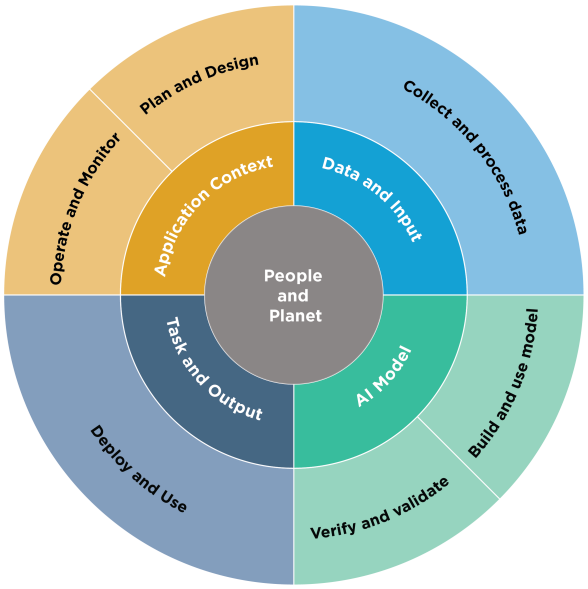


Figure 3.1: Life-cycle and key dimensions of an AI system. (Taken from AI RMF)

Some core aspects of our AI process audit predominantly derives from the guidelines of the United States's NIST[3]. NIST has formulated the AI RMF[4], an AI risk management framework that makes robust recommendations on the

[3] National Institute of Standards and Technology. The official organization of the United States for setting standards for various scientific and technology fields. (NIST)

[4] Artificial Intelligence Risk Management Framework. The framework for risk management that the United States' NIST organization has provided as guidelines. (AI RMF)

governance of AI systems within an organization.

The process comprises the following five broad steps in the overall journey; some steps precede the in-depth technical assessments, and some steps follow thereafter.

- preliminary inventory of the AI landscape
- objectives and strategy review
- technical AI process audit
- stakeholder communication of findings and recommendations
- final assessment report

An overview of the AI audit process from an organizational leadership perspective

The strategic overview of the process

Figure 3.2: A strategic overview of the AI process assessment journey

3.1.1 Preliminary inventory of the AI landscape

This is the beginning of AI process audit: first we familiarize ourselves about the entire AI system and actors within the organization.

GOAL Develop a comprehensive overview of all the AI activities within the organizations, so as to plan the subsequent activities.

The process begins with a preliminary understanding of the following:

1. key business stakeholders
2. landscape of existing and planned AI projects
3. AI engineering & data science teams working with AI
4. data engineering teams that will manage the data that goes into the AI models, and the handling of inferences from these
5. security team
6. AI infrastructure deployments, both on premise as well as those in the various clouds

3.1.2 Objectives and Strategy Review

GOAL Ensure clarity and strong alignment on strategic AI objectives between leadership and the technical teams. [5]

[5] It is rather surprising how often there is a divergence between what the business stakeholders believe the AI initiative will deliver, and what the technical team understand as the objective of their activities! Ensuring a strong alignment early on precludes wasted effort, wasted resources, lost time and subsequent disappointment

Towards this, we perform the following steps:

Business and product objectives: interview the product and business leadership to understand what the specific expectations and objectives are for the AI initiatives, and thus understand their perspective. Do these objectives meet the following criteria:

- realistic or feasible [6]
- well-defined and clearly articulated

[6] Too often, there may be aspects in the objectives that are too expensive to implement, and a slightly different objective may be more sensible or cost efficient.

- measurable or quantifiable in terms of progress and outcomes
- compliant with ethical and regulatory considerations

The ROI of the AI system derives from carefully studying its impact on cost, revenue, derived value and competitive advantage. We ensure that these have been well discussed and understood with respect to the work in progress.

Calculating ROI

Key points for ROI of AI development

EFFECT ON COSTS

- Is the AI product resulting in direct/indirect savings?
- Is the high investment justified by the cost reduction??

EFFECT ON REVENUE

- Has there been a considerable increase in revenue?
- Has it increased the number of customers?

EFFECT ON VALUE

- Are we saving time and/or money?
- Are we improving efficiency?
- Is it worth the effort?

COMPETITIVE ADVANTAGE

- Do we have advantage over the competitors?
- Are we able to differentiate ourselves in the market?

Figure 3.3: Factors in the ROI from AI systems

Technical AI activities and models: interview the AI technical teams (data scientists and AI engineers) to ensure that their activities, and the models under development or created so far, align strongly with the business and product objectives.

Feasibility review: Assess whether the contemporary SOTA[7] methods in AI will allow for the development of AI models that can fulfill the objectives partially or fully. Identify any product requirements which may be infeasible currently, or pose too high of an insufficiently effective AI model.

[7] SOTA: State of the Art

3.1.3 The technical aspects of the AI process audit

The components of the audit of technical aspects of the AI system are rather involved, and each is discussed in great detail in a subsequent section.

3.1.4 Stakeholder communication of findings

This begins the discussion of audit findings, summation, and recommendation of an action plan for remediation where needed. These will be through a draft document, and collaborative meetings with the organization stakeholders from business and engineering.

3.1.5 Final assessment report

After the AI audit, our engineering team will finalize everything into a comprehensive report derived from the draft above, and the subsequent discussions. This report would serve as evidence or certification that an AI process audit has taken place, within the scope of methods described in this document.

3.2 Technical deep-dive into the AI system, processes and practices

Components in the AI audit's technical process

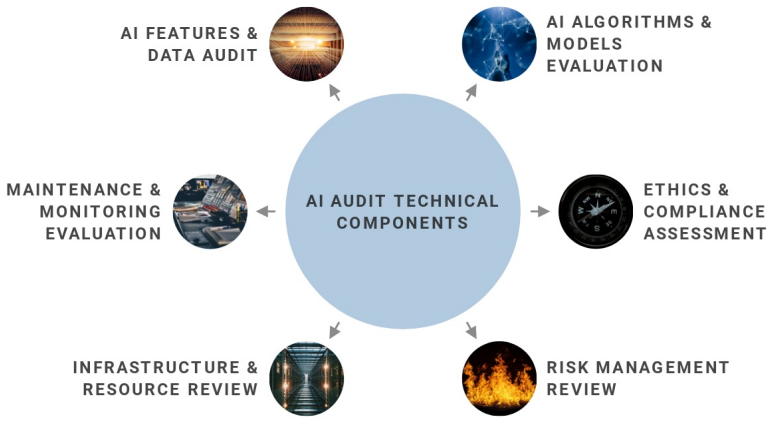


Figure 3.4: Technical components of the AI process audit.

Let us start with a quick overview of what these components involved are:

Data Audit Review the data used in the AI system. Is the data of high quality, relevant, and representative of the problem it's intended to solve? Is the data ethically and legally sourced and managed?

Feature Store Audit See if an robust and rich feature store exists, that feeds into multiple algorithms to accelerate model development and deployment.

Algorithm and Model Evaluation Validate the algorithms and models used by the AI system. Are they fit for purpose? Are they tested and validated with independent data sets?

Performance Assessment Evaluate the performance of the AI system. How accurate are its outputs? How does it handle errors? Are its results interpretable and reliable?

Ethics and Compliance Review Assess the ethical implications and legal compliance of the AI system. Does it protect user data and privacy? Is it fair and unbiased in its decisions? Does it comply with relevant laws and regulations?

Risk Assessment Identify and assess the potential risks of the AI system. This could include operational risks, security risks, reputational risks, and others. Are there appropriate risk mitigation strategies in place?

Infrastructure and Resource Review Examine the AI system's infrastructure and resources. Are they adequate for the system's requirements? Is there a plan in place for scaling up if needed?

Maintenance and Monitoring Evaluation Review the procedures for maintaining and monitoring the AI system. Are there clear protocols for updates and upgrades? How are anomalies and errors detected and resolved?

3.2.1 Details of the technical components of the audit

Now, details follow for each of the components mentioned above.

3.2.2 AI Risk Management

As a powerful technological tool, artificial intelligence brings its unique potential and perils. Since AI adoption is gaining ground at a rapid pace, there is a need for organizations to manage the risks responsibly to prevent adverse impacts. In other words, ensuring that the AI systems are trustworthy and aligned with human goals and values is necessary.

Different geographies have evolving their own set of regulations to manage the risks of AI. In particular, Europe has an Artificial Intelligence Act [8] in a draft form, nearing completion. Europe's AI act requires organizations to measure the risk of their AI systems, and then classify into four varying levels of risk categories. AI systems that pose the highest levels of risk are to be considered as banned. The others need careful evaluation, transparency and audit.

[8]

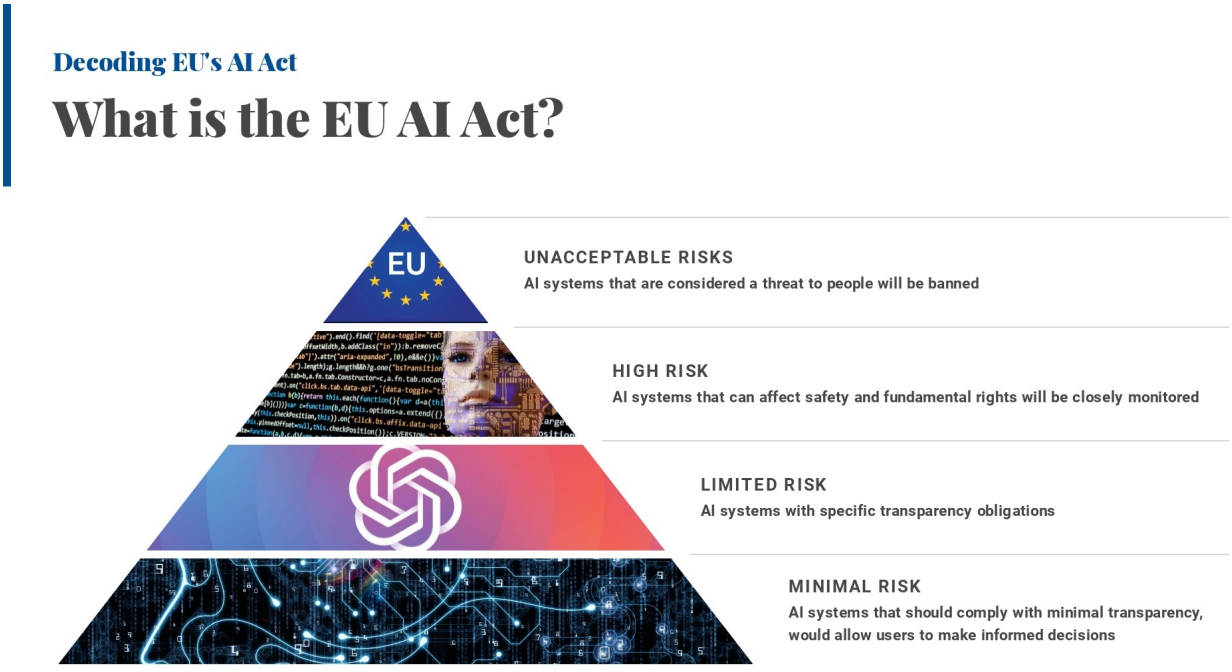


Figure 3.5: The European Artificial Intelligence Act, which categorizes AI applications by levels of risk assessment.

China too has enacted its own set of regulations to shape its AI development.

[9] National Institute of Standards and Technology. The official organization of the United States for setting standards for various scientific and technology fields. (NIST)
[10] Artificial Intelligence Risk Management Framework. The framework for risk management that the United States' NIST organization has provided as guidelines. (AI RMF)
[11]

Close at home, in the United States, the most comprehensive guidelines come from the NIST [9], which has formulated the AI RMF[10].

In our risk audit, we closely adhere to the AI RMF playbook.[11].

It focuses on inculcating an organizational culture of good AI system governance through careful risk management:

Measure Identified risks are assessed, analyzed or tracked.

Manage Risks are prioritized and acted upon based on projected impact.

Map Context and the risk related to it is recognized.

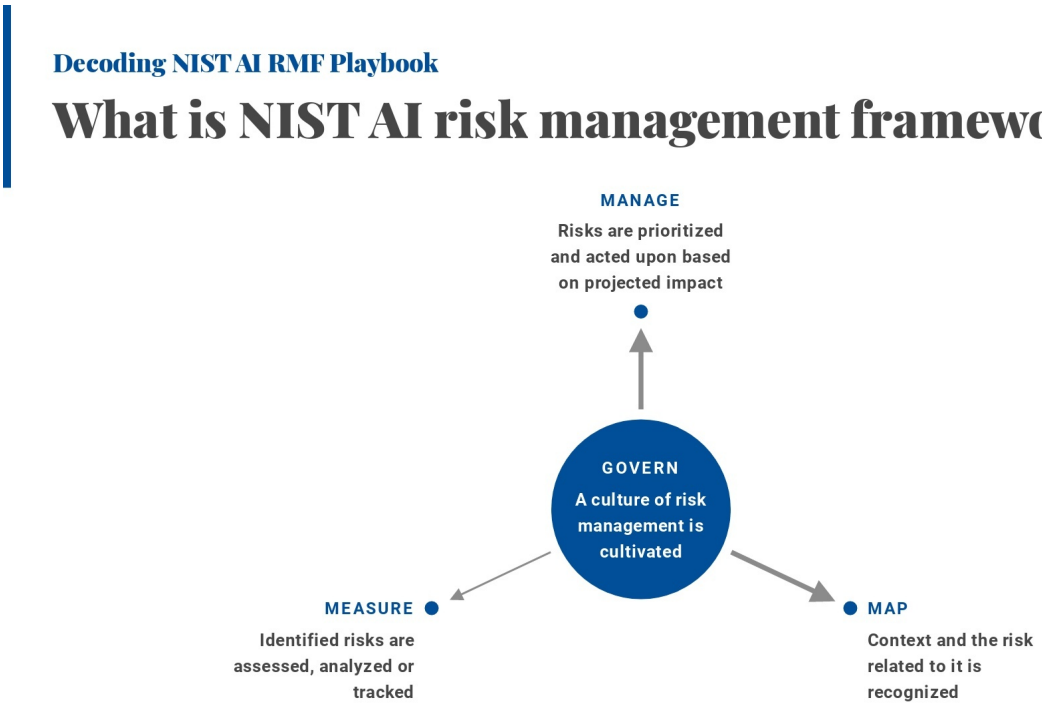
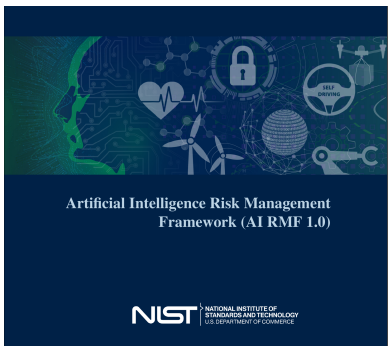


Figure 3.6: NIST Risk Management Framework.



As part of the risk management, the audit goes over the various aspects of the AI RMF playbook, to:

Inform bring more awareness of the many facets and considerations associated with developing and deploying an AI system.

Observe the level of compliance or progress in the various facets.

Recommend activity in areas where work has not begun, but is relevant to the organization.

This follows a detailed approach of test, evaluation, validation and verification of the different aspects of the AI system.

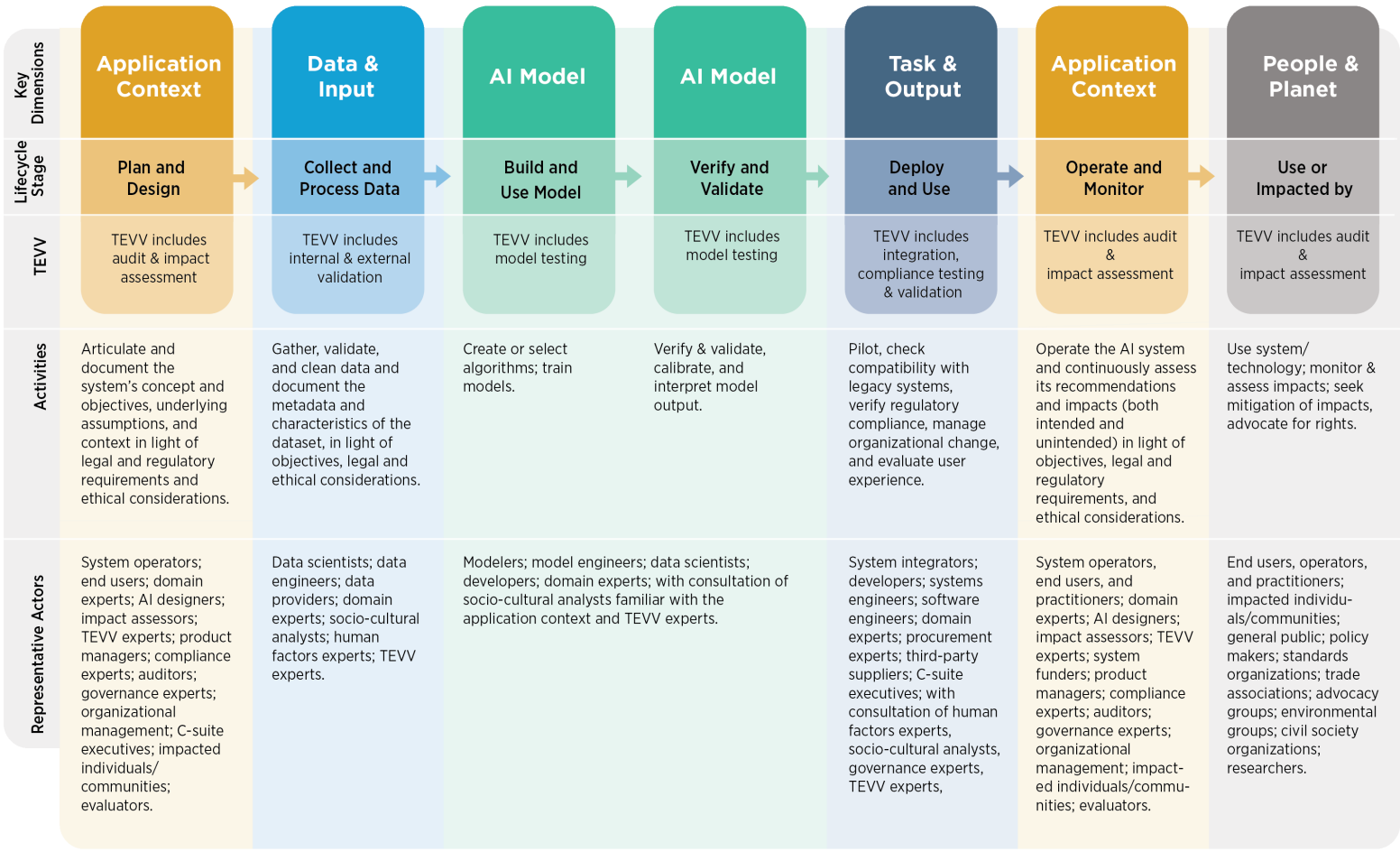


Figure 3.7: Overview of the risk management process, as described by AI RMF

NIST AIRC - Playbook

Type	Title	AI Actors	Topics	Description
Govern	Govern 1.1	Governance and Oversight	Legal and Regulatory, Governance	Legal and regulatory requirements involving AI are understood, managed, and documented.
Govern	Govern 1.2	Governance and Oversight	Trustworthy Characteristics, Governance, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	The characteristics of trustworthy AI are integrated into organizational policies, processes, and procedures.
Govern	Govern 1.3	Governance and Oversight	Risk Tolerance, Governance	Processes and procedures are in place to determine the needed level of risk management activities based on the organization's risk tolerance.
Govern	Govern 1.4	Governance and Oversight	Risk Management, Governance, Documentation	The risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities.
Govern	Govern 1.5	Governance and Oversight, Operation and Monitoring	Continuous monitoring, Governance	Ongoing monitoring and periodic review of the risk management process and its outcomes are planned, organizational roles and responsibilities are clearly defined, including determining the frequency of periodic review.
Govern	Govern 1.6	Governance and Oversight	Risk Management, Governance, Data, Documentation	Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities.
Govern	Govern 1.7	AI Deployment, Operation and Monitoring	Decommission, Governance	Processes and procedures are in place for decommissioning and phasing out of AI systems safely and in a manner that does not increase risks or decrease the organization’s trustworthiness.
Govern	Govern 2.1	Governance and Oversight	Governance, Risk Culture	Roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and are clear to individuals and teams throughout the organization.
Govern	Govern 2.2	Governance and Oversight	Governance, Training	The organization’s personnel and partners receive AI risk management training to enable them to perform their duties and responsibilities consistent with related policies, procedures, and agreements.
Govern	Govern 2.3	Governance and Oversight	Governance, Risk Tolerance	Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.
Govern	Govern 3.1	Governance and Oversight, AI Design	Diversity, Interdisciplinarity, Governance	Decision-makings related to mapping, measuring, and managing AI risks throughout the lifecycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds).
Govern	Govern 3.2	AI Design	Human-AI teaming, Human oversight, Governance	Policies and procedures are in place to define and differentiate roles and

Type	Title	AI Actors	Topics	Description
				responsibilities for human-AI configurations and oversight of AI systems.
Govern	Govern 4.1	AI Design, AI Development, AI Deployment, Operation and Monitoring	Risk Culture, Governance	Organizational policies, and practices are in place to foster a critical thinking and safety-first mindset in the design, development, deployment, and uses of AI systems to minimize negative impacts.
Govern	Govern 4.2	AI Design, AI Development, AI Deployment, Operation and Monitoring	Risk Culture, Governance, Impact Assessment	Organizational teams document the risks and potential impacts of the AI technology they design, develop, deploy, evaluate and use, and communicate about the impacts more broadly.
Govern	Govern 4.3	TEVV, Operation and Monitoring, Governance and Oversight, Fairness and Bias	Risk Culture, Governance, AI Incidents, Impact Assessment, Drift, Fairness and Bias	Organizational practices are in place to enable AI testing, identification of incidents, and information sharing.
Govern	Govern 5.1	AI Design, Governance and Oversight, AI Impact Assessment, Affected Individuals and Communities	Participation, Governance, Impact Assessment	Organizational policies and practices are in place to collect, consider, prioritize, and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks.
Govern	Govern 5.2	AI Impact Assessment, Governance and Oversight, Operation and Monitoring	Participation, Governance, Impact Assessment	Mechanisms are established to enable AI actors to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation.
Govern	Govern 6.1	Third-party entities, Operation and Monitoring, Procurement	Third-party, Legal and Regulatory, Procurement, Supply Chain, Governance	Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third party's intellectual property or other rights.
Govern	Govern 6.2	AI Deployment, TEVV, Operation and Monitoring, Third-party entities	Third-party, Governance, Risk Management, Supply Chain	Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.
Manage	Manage 1.1	AI Deployment, Operation and Monitoring, AI Impact Assessment	AI Deployment, Risk Assessment	A determination is as to whether the AI system achieves its intended purpose and stated objectives and whether its development or deployment should proceed.
Manage	Manage 1.2	AI Deployment, Operation and Monitoring, AI Impact Assessment	Risk Tolerance	Treatment of documented AI risks is prioritized based on impact, likelihood, or available resources or methods.
Manage	Manage 1.3	AI Deployment, Operation and Monitoring, AI Impact Assessment	Legal and Regulatory, Risk Tolerance	Responses to the AI risks deemed high priority as identified by the Map function, are developed, planned, and documented. Risk response options can include mitigating, transferring, avoiding, or accepting.
Manage	Manage 1.4	AI Deployment, Operation and Monitoring, AI Impact Assessment	Risk Response	Negative residual risks (defined as the sum of all unmitigated risks) to both downstream acquirers of AI systems and end users are documented.
Manage	Manage	AI Deployment,	Risk Tolerance, Trade-offs	Resources required to manage AI risks are

Type	Title	AI Actors	Topics	Description
	2.1	Operation and Monitoring, AI Impact Assessment, Governance and Oversight		taken into account, along with viable non-AI alternative systems, approaches, or methods – to reduce the magnitude or likelihood of potential impacts.
Manage	Manage 2.2	AI Deployment, Operation and Monitoring, AI Impact Assessment, Governance and Oversight	AI Deployment, Drift, Societal Values	Mechanisms are in place and applied to sustain the value of deployed AI systems.
Manage	Manage 2.3	AI Deployment, Operation and Monitoring	Risk Response	Procedures are followed to respond to and recover from a previously unknown risk when it is identified.
Manage	Manage 2.4	AI Deployment, Operation and Monitoring, Governance and Oversight	Risk Response, Decommission	Mechanisms are in place and applied, responsibilities are assigned and understood to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.
Manage	Manage 3.1	Third-party entities, Operation and Monitoring, AI Deployment	Third-party, Supply Chain	AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.
Manage	Manage 3.2	Third-party entities, Operation and Monitoring, AI Deployment	Pre-trained models, Monitoring	Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.
Manage	Manage 4.1	AI Deployment, Operation and Monitoring, End-Users, Human Factors, Domain Experts, Affected Individuals and Communities	Monitoring, Participation, AI Deployment, AI Incidents, Risk Response	Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.
Manage	Manage 4.2	TEVV, AI Design, AI Development, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	Monitoring, Impact Assessment, Risk Assessment	Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI actors.
Manage	Manage 4.3	AI Deployment, Operation and Monitoring, End-Users, Human Factors, Domain Experts, Affected Individuals and Communities	AI Incidents, Monitoring	Incidents and errors are communicated to relevant AI actors including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.
Map	Map 1.1		Socio-technical systems, Societal Values, Context of Use, Impact Assessment, TEVV, Trustworthy Characteristics, Validity and Reliability, Safety, Secure	Intended purpose, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented. Considerations include: specific set or

Type	Title	AI Actors	Topics	Description
			and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	types of users along with their expectations; potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet; assumptions and related limitations about AI system purposes; uses and risks across the development or product AI lifecycle; TEVV and system metrics.
Map	Map 1.2		Diversity, Interdisciplinarity, Socio-technical systems	Inter-disciplinary AI actors, competencies, skills and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.
Map	Map 1.3		Socio-technical systems, Societal Values	The organization's mission and relevant goals for the AI technology are understood and documented.
Map	Map 1.4		Context of Use	The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated.
Map	Map 1.5		Risk Tolerance	Organizational risk tolerances are determined and documented.
Map	Map 1.6		Socio-technical systems, Impact Assessment, Documentation	System requirements (e.g., "the system shall respect the privacy of its users") are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks.
Map	Map 2.1		Socio-technical systems	The specific task, and methods used to implement the task, that the AI system will support is defined (e.g., classifiers, generative models, recommenders).
Map	Map 2.2		Limitations, Human oversight, Impact Assessment, Documentation	Information about the AI system's knowledge limits and how system output may be utilized and overseen by humans is documented. Documentation provides sufficient information to assist relevant AI actors when making informed decisions and taking subsequent actions.
Map	Map 2.3	AI Development, TEVV, Domain Experts	TEVV, Data, Impact Assessment, Limitations	Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), system trustworthiness, and construct validation.
Map	Map 3.1	AI Development, AI Deployment, AI Impact Assessment	Socio-technical systems, Documentation	Potential benefits of intended AI system functionality and performance are examined and documented.
Map	Map 3.2	AI Design, AI Development, Operation and Monitoring, AI Design, AI Impact Assessment	Impact Assessment, Trustworthy Characteristics, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability	Potential costs, including non-monetary costs, which result from expected or realized AI errors or system functionality and trustworthiness - as connected to organizational risk tolerance - are examined and documented.

Type	Title	AI Actors	Topics	Description
			and Interpretability, Privacy, Fairness and Bias	
Map	Map 3.3	AI Design, AI Development, Human Factors	Context of Use, Documentation	Targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization.
Map	Map 3.4	AI Design, AI Development, Human Factors, End-Users, Domain Experts, Operation and Monitoring	Human-AI teaming	Processes for operator and practitioner proficiency with AI system performance and trustworthiness – and relevant technical standards and certifications – are defined, assessed and documented.
Map	Map 3.5	Human Factors, End-Users, Domain Experts, Operation and Monitoring, AI Design	Human oversight	Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from GOVERN function.
Map	Map 4.1	Third-party entities, Procurement, Operation and Monitoring, Governance and Oversight	Legal and Regulatory, Third-party, Pre-trained models, Supply Chain, Risk Tolerance	Approaches for mapping AI technology and legal risks of its components – including the use of third-party data or software – are in place, followed, and documented, as are risks of infringement of a third-party's intellectual property or other rights.
Map	Map 4.2	AI Deployment, TEVV, Operation and Monitoring, Third-party entities	Third-party, Pre-trained models	Internal risk controls for components of the AI system including third-party AI technologies are identified and documented.
Map	Map 5.1	AI Design, AI Development, AI Deployment, AI Impact Assessment, Operation and Monitoring, Affected Individuals and Communities, End-Users	Participation, Impact Assessment	Likelihood and magnitude of each identified impact (both potentially beneficial and harmful) based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system, or other data are identified and documented.
Map	Map 5.2	AI Design, Human Factors, AI Deployment, AI Impact Assessment, Operation and Monitoring, Domain Experts, Affected Individuals and Communities, End-Users	Participation, Impact Assessment	Practices and personnel for supporting regular engagement with relevant AI actors and integrating feedback about positive, negative, and unanticipated impacts are in place and documented.
Measure	Measure 1.1	AI Development, TEVV, Domain Experts	Trustworthy Characteristics, Risk Assessment, TEVV, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	Approaches and metrics for measurement of AI risks enumerated during the Map function are selected for implementation starting with the most significant AI risks. The risks or trustworthiness characteristics that will not – or cannot – be measured are properly documented.
Measure	Measure 1.2	TEVV, AI Impact Assessment, AI Development, AI	Impact Assessment, TEVV, Context of Use	Appropriateness of AI metrics and effectiveness of existing controls is regularly assessed and updated including

Type	Title	AI Actors	Topics	Description
		Deployment, Affected Individuals and Communities		reports of errors and impacts on affected communities.
Measure	Measure 1.3	TEVV, AI Impact Assessment, AI Development, AI Deployment, Affected Individuals and Communities, Domain Experts, End-Users, Operation and Monitoring	Participation, Impact Assessment, Context of Use	Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, AI actors external to the team that developed or deployed the AI system, and affected communities are consulted in support of assessments as necessary per organizational risk tolerance.
Measure	Measure 2.1	TEVV	TEVV, Documentation, Validity and Reliability	Test sets, metrics, and details about the tools used during test, evaluation, validation, and verification (TEVV) are documented.
Measure	Measure 2.2	TEVV, Human Factors, AI Development	Data, Human Subjects Protection	Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population.
Measure	Measure 2.3	TEVV, AI Deployment	TEVV, Impact Assessment	AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s). Measures are documented.
Measure	Measure 2.4	AI Deployment, TEVV	TEVV, Monitoring, Drift	The functionality and behavior of the AI system and its components – as identified in the MAP function – are monitored when in production.
Measure	Measure 2.5	TEVV, Domain Experts	TEVV, Validity and Reliability, Trustworthy Characteristics, Data	The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.
Measure	Measure 2.6	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Safety, Trustworthy Characteristics, Context of Use	AI system is evaluated regularly for safety risks – as identified in the MAP function. The AI system to be deployed is demonstrated to be safe, its residual negative risk does not exceed the risk tolerance, and can fail safely, particularly if made to operate beyond its knowledge limits. Safety metrics implicate system reliability and robustness, real-time monitoring, and response times for AI system failures.
Measure	Measure 2.7	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Secure and Resilient, Trustworthy Characteristics	AI system security and resilience – as identified in the MAP function – are evaluated and documented.
Measure	Measure 2.8	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Transparency and Accountability, Trustworthy Characteristics	Risks associated with transparency and accountability – as identified in the MAP function – are examined and documented.
Measure	Measure	TEVV, Domain	TEVV, Explainability and	The AI model is explained, validated, and

Type	Title	AI Actors	Topics	Description
	2.9	Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users	Interpretability, Trustworthy Characteristics	documented, and AI system output is interpreted within its context – as identified in the MAP function – and to inform responsible use and governance.
Measure	Measure 2.10	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users	TEVV, Privacy, Trustworthy Characteristics	Privacy risk of the AI system – as identified in the MAP function – is examined and documented.
Measure	Measure 2.11	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users, Affected Individuals and Communities	TEVV, Fairness and Bias, Trustworthy Characteristics	Fairness and bias – as identified in the MAP function – is evaluated and results are documented.
Measure	Measure 2.12	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Environmental Impact	Environmental impact and sustainability of AI model training and management activities – as identified in the MAP function – are assessed and documented.
Measure	Measure 2.13	TEVV, AI Deployment, Operation and Monitoring	TEVV, Effectiveness	Effectiveness of the employed TEVV metrics and processes in the MEASURE function are evaluated and documented.
Measure	Measure 3.1	TEVV, AI Impact Assessment, Operation and Monitoring	TEVV, Monitoring, Continual Improvement	Approaches, personnel, and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks based on factors such as intended and actual performance in deployed contexts.
Measure	Measure 3.2	TEVV, Domain Experts, AI Impact Assessment, Operation and Monitoring	Monitoring	Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available.
Measure	Measure 3.3	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	Participation, Contestability, TEVV, Impact Assessment	Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.
Measure	Measure 4.1	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	TEVV, Participation, Context of Use	Measurement approaches for identifying AI risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.
Measure	Measure 4.2	TEVV, AI Deployment, Domain Experts,	TEVV, Participation, Trustworthy Characteristics, Validity and Reliability,	Measurement results regarding AI system trustworthiness in deployment context(s) and across AI lifecycle are informed by

Type	Title	AI Actors	Topics	Description
		Operation and Monitoring, End-Users	Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	input from domain experts and other relevant AI actors to validate whether the system is performing consistently as intended. Results are documented.
Measure	Measure 4.3	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	TEVV, Participation, Trustworthy Characteristics, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	Measurable performance improvements or declines based on consultations with relevant AI actors including affected communities, and field data about context-relevant risks and trustworthiness characteristics, are identified and documented.

3.2.3 AI Data Audit

There are the following aspects to study in the context of data that flows into AI model training:

Data governance and quality As the dictum, garbage-in, garbage-out implies,a system that relies on data will yield degraded performance in the presence of noisy or poorly maintained data. This is particularly so for AI models, which inherently learn from the data, and consequently, their training gets adversely affected with poor data quality.

We need to carefully check the subset of data that flows into the creation of machine learning features for model training for such pathologies as:

- significantly high proportion of noisy or junk data instances in the dataset, that would adversely affect training
- presence of anomalies and outliers that are not carefully handled or documented for. [12]
- excessive missing values in crucial rows to the point to rendering these features ineffective in the model training
- a lack of missing values treatment and inputation strategy for the features
- presence of statistically irrevelant columns or attributes such as database row keys marked as features in model training
- non-informative data values, that the model may inadvertently consider a learnable signal. [13]

[12] Presence of anomalies is not necessarily bad always – indeed, if the purpose of a particular model is anomaly detection, then anomalies are a necessary part of the training dataset.

[13] For example, one may find in a dataset comprising of `[image, caption]` tuples many captions that are non-informative text such as "image", "qwert", "asdfg", "1234", "?", or simply stuffed with irrelevant keywords for SEO.

Data privacy Preserving the privacy of data is both an ethical and regulatory imperative. The problematic aspects in the data that needs investigation are:

- presence of overt PII[14] such as social security number, name, etc. that not only have no statistical relevance to AI modeling, but potentially lead to serious data privacy issues
- data containing demographic information, such as gender, age, race, etc. In most situations, unless there is a strong justification for using them on a case by case basis, these should have been filtered out
- data from which some approximation to a personally identifiable information may be inferred.[15]

[14] Personally identifiable information. Any data that directly or indirectly leads to the identification of a person is considered PII (PII)

[15] There are measures such as k-anonymity, l-diversity and t-closeness that must be considered to ensure sufficient anonymity and privacy is preserved.

Data insufficiency There may be insufficient dataset to train a complex model, such that there is significant hazard of model over fitting, leading to misleading or erroneous predictions

Data aging and obsolescence the data may be too old or rendered obsolete over time, and thus not representative of the current data over which the AI model needs to make prediction. Thus the training dataset and the runtime inference dataset in production will then now have the same underlying probability distributions, violating basic assumptions needed to build an effective AI model.

DLP AI systems presume robust DLP enforcement beyond simply privacy considerations. Weak or absent practices in this context is a significant red flag that needs specific and urgent remediation. In particular, this audit will look for and document the existing tools and safeguards in this context, and evaluate their effectiveness. [16]

[16] Data Loss Prevention. It refers to organizations processes, safeguards and software tools that ensure that sensisitive data is not lost, damaged, misused, stolen or accessed by unauthorized users. Having a robust DLP enforcement is often a regulatory requirement for HIPAA, GDPR, PCI-DSS, SOC-2, FedRAMP, and other data privacy rules specific to various nations. (DLP)

3.2.4 Tools available or used

There are a plethora of mature tools available for DLP enforcement and each organization makes its own particular choices; in this audit we ensure that tool chain comprises a robust collection that covers all aspects, and provides sufficient safeguards.

Popular PII detection tools

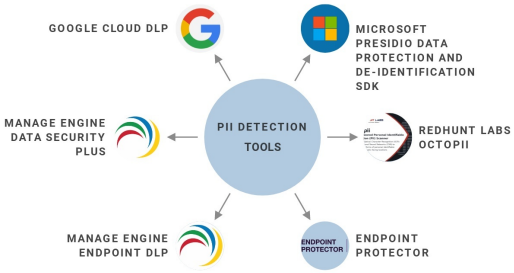


Figure 3.8: A sample of widely used DLP enforcement tools.

We pay particular attention to the preservation of PII. We will take some of the data entering AI systems, and check to ensure that they contain no traces of personally identifiable information.

For example, we often use `Presidio`, a python library to scan the data. [17]

[17] Needless to say, we ensure that the data never leaves the secure intranet of the organization: we deploy the tools there in the data-residency zone, if such tooling does not already exist.

Microsoft's Presidio model is an open source library designed specifically for detecting PII in unstructured data. By scanning the data, we can identify any sensitive information present, such as social security numbers, credit card numbers, addresses, etc. See full list here[18].

[18] What does it do?

Presidio leverages machine learning algorithms to recognize patterns and entities commonly associated with PII. Some of the advantages include:

- predefined and custom Named Entity Recognition, regular expressions, rule-based logic, and checksum across multiple languages.
- connectivity to external PII detection models.
- multiple usage options - Python, PySpark, Docker, Kubernetes support.
- customizable PII identification and anonymization.
- redaction module for PII text in images.

How the audit is done

To perform the scanning, we follow these steps:

- obtain the data dump file from the company's designated source.
- install and configure[19] Microsoft's Presidio model.
- determine the PII categories to detect, or create custom categories.
- write Python code to scan the data.
- execute the scanning code to identify potential PII.

[19] Installation steps: https://microsoft.github.io/presidio/installation/

As such our pipeline of audit would be as follows:

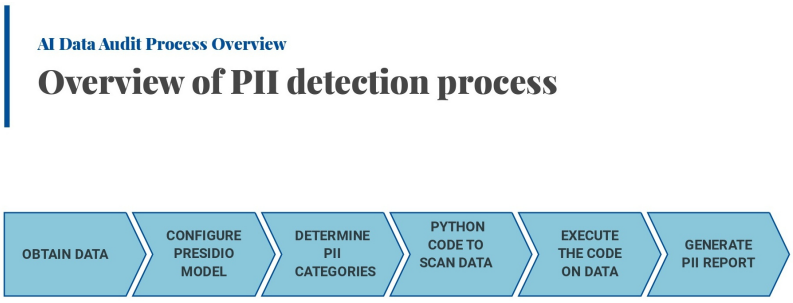


Figure 3.9: A data scanning for PII violations, using the Python programming library, `Presideo`.

Towards this end, we ask the engineers to take a dump of their data, and run the tool directly, producing the report.

Presidio scanning can reveal instances of PII present in the company data. It will:

Results of the investigation

- classify the detected PII into different categories, such as names, addresses, phone numbers, social security numbers etc.[20]
- Provide a report highlighting the location (start and end positions) and type of sensitive information found.

[20]

3.2.5 AI code scan for violations

There AI code-base development often have are a few pathologies, that can degrade the effectiveness, performance and reliability of an AI system. In particular, some common pathologies are:

Security violations There may be serious security vulnerabilities in the AI system code-base. It may either be in the inference serving code, the micro-services, or the AI models themselves.

Common occurring defects While some minor defects in the code are inevitable, it becomes a cause of worry when there is a preponderance of unaddressed critical and major defects in the code-base.

Weak or non-existence test-suite A weak or non-existent test-suite is a significant red flag: while the business exigencies of needing fast releases and delivery to market may temporarily drive a hasty development practice, a complete absence of test-suite and poor test coverage points to a more systemic problem with the engineering development culture.

Technical debt The overall technical debt may have accumulated to an unacceptable amount, to the extent that new product development is impeded with the instabilities in existing AI system, and the continual engagement of resources to stabilize it in production.

In this audit, we specifically focus on, and limit the scope of our investigations

to the code base pertaining to the AI systems. This includes studying the code-base associated with the following:

AI models architecture The actual AI model training and architecture.

Model inferences The serving of the model inferences at runtime.

Feature extraction The extraction of feature vectors from the underlying Data-lake.

Data manipulation The various steps of data extractions, and transformations that lead to the eventual features for AI modeling.

Data cleaning The process of cleansing the raw data of its avoidable noise.

Missing value handling This includes the missing-values analysis and handling, and methods for imputation of values.

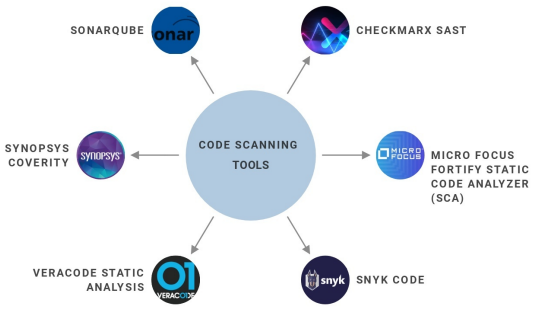
[21] Exploratory data analysis. (EDA)

EDA EDA[21] is a significant step in the modeling process. We ensure that this has been given extensive and appropriate coverage in the code base, wherever necessary.

MLOps The code-base and scripts associated with the MLOps.

Figure 3.10: Some common static code analysis tools

Popular code scanning tools



For static code analysis, there are many mature tools in existence, and we investigate and evaluate the deployment and use of these in the organization for the AI systems.

If any such tools exist, we will investigate the findings therein, and provide our evaluation. If one does not exist, we tend to prefer SonarQube[22] as our static code analyzer of choice, and thus will deploy it to extract the analysis results.

[22] A popular, open-source code quality monitoring dashboard. It helps to identify and resolve code quality and security issues in software projects. By scanning the codebase, we can identify potential vulnerabilities, bugs, code smells and

AI Code Audit Process Overview

Overview of code scanning process

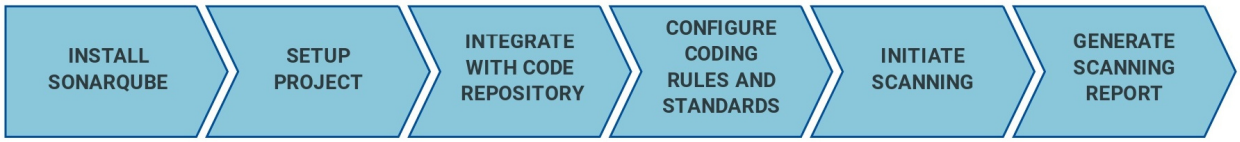


Figure 3.11: Static code analysis pipeline during audit.

To elaborate more on this, let us assume that we are using SonarQube. It performs an in-depth analysis of the code, checking for adherence to coding standards, potential security vulnerabilities, and code smells.

23

[23] Either on-premise deployment of SonarQube or a cloud deployment of SonarCloud may be best suited for an organization's needs, on a case-by-case basis, for the purposes of this audit.
How the code scan occurs

To conduct the code scanning, we can follow these steps:

- install and set up SonarQube in the company's infrastructure.
- setup SonarQube project and integrate it with Python code repository.
- configure SonarQube to use appropriate coding rules and standards.
- initiate the scanning process to evaluate the code-base for potential issues.

More details are available at the official SonarQube documentation portal.

The features of SonarQube that we focus on help us to:

- identify critical and major code vulnerabilities that need immediate attention
- alert coding issues that impact code readability and maintainability (extensive copy-paste, etc.)
- provide an overall code quality score along with a detailed report on the identified issues
- assess security vulnerabilities and code smells

Once we have a full report of the static code analysis, we apply experience and judgment to create a plan for the needed remediation of the most critical violations.

Some baseline recommendations

More broadly, we recommend and check for compliance with the following best practices:

- PII data protection with a mature and robust set of DLP tools
- implement data protection policies to handle data
- encrypt sensitive data, at rest and in transit.
- monitor and audit access to PII data, and prevent unauthorized use
- develop a comprehensive data handling policy that defines how data is collected, stored, processed, and disposed of properly
- conduct regular data protection awareness and security training for employees
- periodically scan the data to ensure DLP compliance and maintain data security

3.2.6 Evaluation of the MLOps Infrastructure and data pipelines

Need to work on this section next.

3.2.7 Strictly restricted access to the AI inference servers

Need to work on this section next.

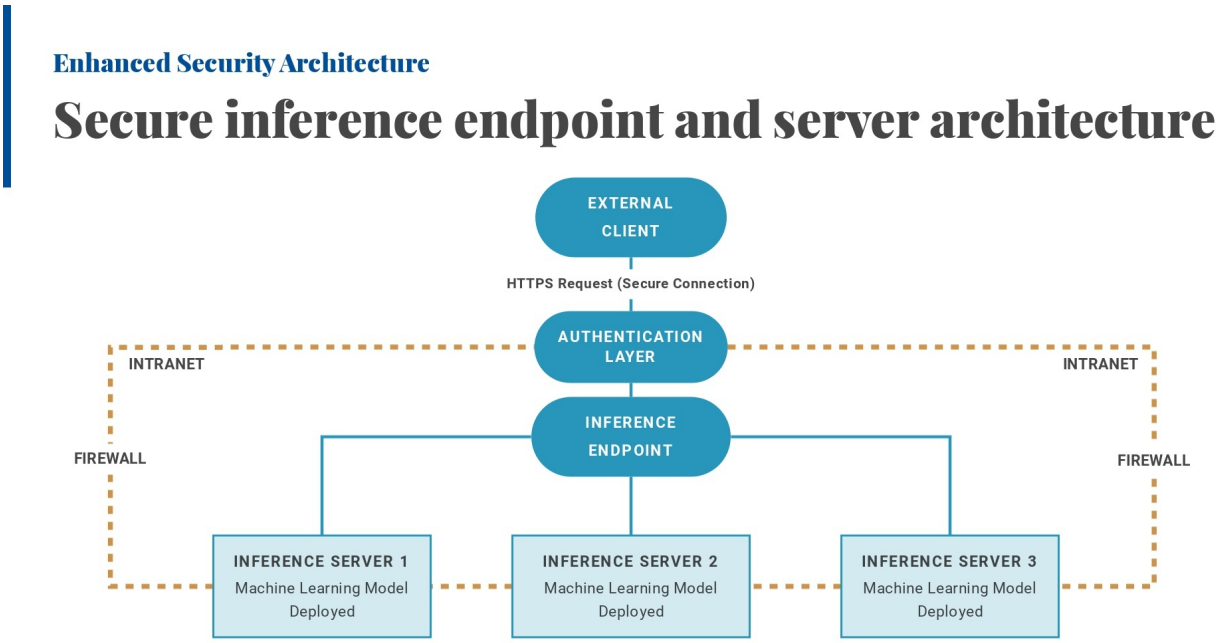


Figure 3.12: Ensuring that the AI inference servers are always protected behind multiple firewalls, and exist in a vlan of its own with only authenticated access from trusted applications within the intranet.

ai audit proc

Glossary

Notation	Description	Page List
AI RMF	Artificial Intelligence Risk Management Framework. The framework for risk management that the United States' NIST organization has provided as guidelines.	12, 20, 22, 41
Anonymization	Anonymization is a process of removing or replacing personal identifiable information in a data sets, in order to create anonymity	41
BigTable	Bigtable is a compressed, high performance, proprietary data storage system built on Google File System. Bigtable supports heavy read and writes.	41

s audit

Notation	Description	Page List
CICD	Continuous Integration and Delivery: the practice of continually integrating and testing from the head of the source-control capturing incremental changes, and then doing a continual deployment process, leading to frequent, micro-releases rather than large releases after a long gestation period.	41
CII	Any data that directly or indirectly leads to the identification of a person is considered CII	41
Cohort	A cohort is a group of subjects who form the context of a study or analysis	41
Columnar Database	A columnar database is column-oriented database management system that stores data tables by columns rather than by row,which makes it suitable for analytical query processing, and thus for data warehouses.	41
Compliance	Compliance is the conformance to a set of rules or regulations.	41
Container	A container is an atomic unit of software or application that packages all the relevant binaries, configuration, and all their dependencies. This allows for easy deployment, migration across environments and machines, continuous delivery of releases, and facilitates scale-out of the application under heavy traffic.	41

Notation	Description	Page List
Data-lake	A data lake is usually an organization-wide, consolidated big-data store where all forms of data – structured as well as unstructured data – are kept. This data therefore can be the target of various data wrangling activities, multidimensional analysis and report generation, as well as be the fodder for various machine-learning/AI pipelines for the extraction of patterns, or building of predictive models.	36, 41
DLP	Data Loss Prevention. It refers to organizations processes, safeguards and software tools that ensure that sensisitive data is not lost, damaged, misused, stolen or accessed by unauthorized users. Having a robust DLP enforcement is often a regulatory requirement for HIPAA, GDPR, PCI-DSS, SOC-2, FedRAMP, and other data privacy rules specific to various nations.	33, 38, 41
EDA	Exploratory data analysis.	36, 41
Kubeflow	Provides many of the MLOps functionality when deploying to Kubernetes clusters. https://kubeflow.org	41
LLM	Large Language Model. These are also called large transformers. These are crucial to our machine learning tasks, and we will use these for semantic text and image embeddings, question-answering, and other tasks in the recommenders.	41

Notation	Description	Page List
MLOps	The continuous build and delivery automation pipeline in machine learning, which also includes the model evaluation for metrics, and tracking of improvements.	41, 46
MLP	In our context, it refers to the overall Machine Learning Platform within a given organization. It is the architecture and scalable statistical and machine-learning platform for a comprehensive set of components, from the big-data lake, computational layer, multidimensional cubes, feature store, MLOps, data Pipeline, and the AI model training and inferences.	41
NIST	National Institute of Standards and Technology. The official organization of the United States for setting standards for various scientific and technology fields.	12, 20, 41, 43
OECD	The Organisation for Economic Co-operation and Development. It is an international organization that aims to create better policies that positively impact lives by fostering prosperity, equality, opportunity and well-being for all.	12, 41, 46
OECD.AI	OECD's policy observatory for artificial intelligence.	41
PII	Personally identifiable information. Any data that directly or indirectly leads to the identification of a person is considered PII	32–35, 38, 41

Notation	Description	Page List
Pipeline	Pipeline is a set of data processing unit connected in series where the output of one unit is the input to the next one.	41, 46
Runbook	Runbook: this is a technical document that describes the exact processes, which if followed, lead to achieving an objective. An example is a runbook for data ingestion. A runbook here helps ensure that any available resource in the team can follow it, and thus perform the weekly data ingestion process.	41
SonarQube	A popular, open-source code quality monitoring dashboard. It helps to identify and resolve code quality and security issues in software projects. By scanning the codebase, we can identify potential vulnerabilities, bugs, code smells and maintainability issues.	
https://docs.sonarqube.org/latest/		37, 41
SOP	Standard Operating Procedure: a document that articulates the operational procedures associated with a specific component.	41
SOTA	State of the art. It implies that a particular implementation aligns with the best comtemporary benchmarks, standards and practices.	12, 41
SupportVectors	SupportVectors AI Lab is a silicon valley tech company that primarily focuses on crafting and implementing the architecture for AI-centered cloud application platforms, and building AI machine learning models.	41

s audit

Notation	Description	Page List
TEVV	Test and evaluation, validation and verification.	41

Bibliography

[NIS23] NIST. “NIST AI RMF Playbook”. In: (2023). url: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

[OEC22] OECD. “OECD Framework for the Classification of AI systems”. In: 323 (2022). doi: https://doi.org/https://doi.org/10.1787/cb6d9eca-en. url: https://www.oecd-ilibrary.org/content/paper/cb6d9eca-en.

[] OECD Policy Observatory. url: https://oecd.ai/en/.

[Pre] Microsoft Presidio. PII entities supported by Presidio. url: https://microsoft.github.io/presidio/supported_entities/.

[Sio21] Lucilla Sioli. A European Strategy for Artificial Intelligence. 2021. url: https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf.

[Son] SonarQube. SonarQube official documentation portal. url: https://docs.sonarsource.com/sonarqube/latest.

[Uni] The European Union. The Artificial Intelligence Act (Europe). url: https://artificialintelligenceact.eu/.

[UT23] National Institute of Standards United States NIST and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023.

There were 51 pages in this document.

FEEDBACK

Please send any feedback on this document, or report errors to the author at asif@supportvectors.com