

Navigating Uncharted AI Waters through Optimized Governance

Aligning AI initiatives with strategic goals and regulatory requirements

TABLE OF CONTENTS

Introduction	3
Understanding AI Governance	4
Categorizing AI Systems by Risk and Utility	5
Role of Organization, AI Provider versus AI Deployer (Under the AI Act):	6
Risk Management	7
Key Risk Factor 1: Risk of AI Deployer becoming the AI Provider (Due to substantial modification):	7
Key Risk Factor 2: Model Training - Continuously Retrained versus Batch Training	8
Key Risk Factor 3: Crossing into Prohibited Practices (Article 5).....	8
Key Risk Factor 4: Lack of Transparency	9
Key Risk Factor 5: Data Quality and Bias in AI Systems	9
Key Risk Factor 6: Insufficient Human Oversight	10
Risk Evaluation & Mitigation.....	11
Developing & Implementing Your AI Governance Framework	12
Establishing an Organizational AI Governance Structure	12
Setting Clear Objectives	13
Establishing/Extending Policies and Procedures	13
Practical First Steps for AI Governance	13
Quick Wins	14
Selecting the Right AI Vendor Partner	15
Criteria for Evaluating AI Vendors.....	16
Artificial Intelligence Screening Questions for Suppliers: Guard-rails on the Road to AI Assistance	17
Conclusion	18
Appendices	19
Glossary of Key Terms: EU AI Act.....	19
Our Top 3 AI Governance Resources.....	23



INTRODUCTION

The echoes of technological innovation reverberate through time, shaping the world we live in today. In the late 1970s, artificial intelligence (AI) made humble beginnings with basic voice recognition systems capable of distinguishing just ten simple words. Fast forward to the present, and AI technologies have evolved exponentially. They not only recognize voices but can also replicate them with astonishing accuracy, mimicking an individual's phrasing and tone. This leap in capability brings significant benefits, such as enhanced customer experiences and innovative business models. However, it also introduces risks like unauthorized impersonation, deepfake fraud, and privacy breaches.

For instance, a notable case of the misuse of deepfake technology to create realistic but fake audio and video content saw a finance employee at a multinational firm deceived into transferring \$25 million to fraudsters who impersonated the company's CEO during a video conference call. The employee believed he was interacting with other staff members, but all participants were deepfake recreations. This highlights the dual nature of AI advancements: the potential for innovation and the threat of misuse. The swift progression of AI prompts crucial questions about how organizations can navigate challenges and opportunities, especially regarding governance and ethical considerations. As AI reshapes various

aspects of society, from personalized healthcare to autonomous vehicles, harnessing potential while safeguarding against misuse has never been more important.

Emerging compliance issues and forthcoming legal changes related to copyright and intellectual property underscore the necessity for companies to design, build, and operate AI technologies that are ethical, human-centered, and trustworthy. Business managers must proactively assess the impact of AI on customers and employees, implementing robust policies to minimize risks and prevent harm. Developing an AI governance framework and adopting best practices like ethics by design enable businesses to promote the responsible creation and use of artificial intelligence.

A recommended dual approach involves establishing an in-house AI governance function while ensuring that critical vendors also practice good governance. Cross-functional collaboration, engaging legal, technical, and operational teams, is key to creating a comprehensive governance strategy that aligns with both business objectives and societal expectations.

With nearly a quarter of a century of experience in developing AI-based fraud prevention solutions across multiple industries worldwide, Daon offers extensive knowledge in AI governance. This book aims to assist businesses in implementing AI solutions responsibly by sharing insights and recommendations on approaching governance. Through practical examples and actionable steps, we hope to guide your organization toward the next important steps in responsible and trustworthy AI development and deployment.

UNDERSTANDING AI GOVERNANCE

Artificial intelligence (AI) is an umbrella term encompassing a wide array of technologies, from basic automation tools and machine learning algorithms to advanced systems designed to incorporate layers with multiple, interlinked AI agents. As a new frontier in business processes, AI offers unprecedented opportunities alongside significant challenges. The level of risk and oversight required for AI systems differs based on their function, capabilities, and potential impact on society.

This variability is reflected in regulations like the EU AI Act, which adopts a risk-based approach to governing AI technologies. Although this book does not exclusively focus on the EU AI Act, we do use this Act as a framework to discuss AI governance. The Act categorizes AI systems into different risk levels: minimal, limited, high, and unacceptable based on their potential to impact health, safety, and fundamental rights. High-risk AI systems, such as those used in critical infrastructure, education, employment, and law enforcement, are subject to stringent requirements. AI Providers (the entity putting the system on the market e.g. the developer) must meet obligations including comprehensive documentation, transparency measures, performance metrics, and a post-market monitoring plan. These requirements ensure that AI Deployers (the users of the technology) can implement the systems responsibly.

The AI Act follows other European regulations like the General Data Protection Regulation (GDPR), which informed the design of frameworks from organizations such as ISO and NIST to provide specific guidelines for the governance of technology systems in organizations. These frameworks emphasize data privacy, security, and ethical considerations in line with global legislation, reinforcing the need for organizations to adopt robust governance practices.

It's important to recognize that AI systems do not "think" in the human sense; they are designed to perform specific tasks based on the data they have been trained on. However, with sufficient training data, AI can produce exceptional outputs, sometimes surpassing human capabilities in certain domains. This is particularly evident in large language models (LLMs) trained on large datasets such as OpenAI's ChatGPT which generate impressive human-like outputs. It should be noted that outputs from LLMs come with risks such as data leakage, where sensitive data used to train the model can accidentally be provided to end users, and hallucinations, where the model provides seemingly coherent responses, but are entirely fabricated facts.

Machine learning (ML) models and LLMs can be extended by programming individual Generative Pre-training Transformers (GPTs), or ML algorithms to form complex multi-agent systems where multiple GPTs or algorithms collaborate on various tasks, forming advanced AI systems.

The type of oversight and transparency needed varies significantly across different AI applications. For instance, a simple ML algorithm used to generate insurance quotes requires different governance measures compared to a multi-agent system utilizing agents trained on large language models. The latter involves more complex decision-making processes and can generate human-like text, increasing the need for oversight at multiple stages of the system. As a less complex system, the design of human oversight requirements for a machine learning model is, more straightforward. Despite the differences in oversight requirements, both systems have ethical and compliance considerations to be implemented from system design, data collection, model training, system deployment, and ongoing operation of the system.

Many AI systems operate using “black box” models, where the internal decision-making processes are not fully explainable even to their developers. This lack of transparency poses challenges for accountability and trust. AI systems that utilize continuous learning, i.e. updating their models in real-time as new data becomes available, can be unpredictable as they are constantly evolving. Their dynamic nature means their impact on business operations can be far-reaching and potentially volatile.

When exploring the application of AI in your organization, it’s crucial to understand the types of AI systems involved and develop a tailored governance plan for your technology. Understanding the nuances of AI governance is essential for mitigating risks and leveraging the full potential of AI technologies. This book gives guidance and practical steps to establish a governance plan to enable your organization to proactively address risks, ethical considerations, foster trust, ensure compliance, and drive innovation responsibly.

Categorizing AI Systems by Risk and Utility

In the EU model, risk is based on the function of the system. Even a straightforward ML model, if used for any of the functions identified as high-risk, must adhere to all the legal requirements of high-risk systems.

A simplified breakdown of those functions includes:



Biometrics

Any biometric analysis that takes place outside of the individual’s direct knowledge, especially for tasks like categorization or emotional recognition.



Critical infrastructure

Providing for general safety or basic utilities.



Education/ vocational training

Any systems used to provide access, assessment, or monitoring in an educational environment.



Employment

Any systems used for recruiting, assessment, work allocation, or monitoring in a business environment.



Service access

Systems designed to determine eligibility for public or private services, including credit scores, and for dispatching emergency services.



Law enforcement

Any systems used by or on behalf of any law enforcement agency, especially to assess risk or evaluate individuals or evidence.



Immigration/ border control

Systems used to determine eligibility and risk, both security and health, for border access, as well as determining nationality outside of verifying travel documents.



Justice/Elections

Systems used to interpret or apply law or in any way influence political campaigns.

The primary focus when determining high risk is if it relates to health, safety, and fundamental rights. However, when establishing your own internal risk classification, you should also evaluate the risk to your business. One recommended approach is to classify each AI system and establish internal governance protocols for each classification group. Some examples of classifications, organized by risk, function, and training, are proposed in the table below. These can be used individually, in combination, or be expanded upon to fit your organization. Key risk areas within your systems and ensuring that you have the human oversight and risk management process in place for these systems should be your focus.



**Classification 1
RISK LEVEL**

- Minimal Risk
- Limited Risk
- High-Risk
- Prohibited Practices



**Classification 2
TRAINING**

- Machine Learning (supervised, semi & un-supervised)
- Hybrid learning (neural systems, conversational AI, continuously trained)



**Classification 3
FUNCTION**

- Administration & Optimization
- Customer Facing
- Decision Making/Support
- Strategic Planning
- Innovation & New Products



**Role of Organization,
AI Provider versus AI Deployer (Under the AI Act):**

While there are a number of possible roles that a business can take, there are two key roles to focus on when establishing your organization’s AI governance responsibilities. The AI Act refers to these as the AI Provider and AI Deployer, and failure to comply with the responsibilities stipulated for each role results in significant fines. Article 5, which covers prohibited practices such as manipulation affecting or exploiting individuals or groups in a way that causes them significant harm, incurs the highest fines of up to €35,000,000 or 7% of annual worldwide turnover, whichever figure is higher. Breaches of the remainder of the act, still carry

significant fines of up to €15,000,000 or 3% of annual worldwide turnover. For this reason, it is vital to establish the role of your organization for each AI system in order to properly assess the responsibilities that must be met.

As defined by the AI Act, AI Providers are the entities who place the high-risk AI system on the market, for example, a software developer, and are responsible for the bulk of documentation and ethical requirements. They must design, build, and offer the ability to monitor the AI system in a way that satisfies the AI Act’s requirements. These requirements vary by function of the AI system, but compliance will largely be met by adhering to the Seven Principles of Trustworthy AI outlined in the AI Act.

The **Seven Principles** of Trustworthy AI

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination, and fairness
6. Societal and environmental well-being
7. Accountability

The AI Act defines the AI Deployer as the professional entity that uses the AI system. The AI Deployer's responsibilities include implementing the post-monitoring plan supplied by the AI Provider, ensuring transparency and accountability of their systems, and having risk management processes in place. It is crucial that the AI Deployer understands what is expected from their AI Provider and implements the system as it was intended to be implemented.

Once you have identified the AI systems you are

RISK MANAGEMENT

using, and your role in relation to those systems, you should establish the level of risk of these systems to your organization. The AI Act states that an organization's risk management system should be a continuous, iterative process that is planned and implemented throughout the entire lifecycle of a high-risk AI system. This process should focus on identifying and mitigating relevant risks that artificial intelligence systems may pose to health, safety, and fundamental rights. This section outlines some key considerations related to the risk of these systems and provides a table to assist with a simple classification of these for a specific case study in the financial sector.

Key Risk Factor 1:

Risk of AI Deployer becoming the AI Provider (Due to substantial modification):

If an AI Deployer changes the way that the system they are using from a third-party AI Provider is implemented, they are at risk of becoming the AI Provider of that system, making them responsible for the requirements of the AI Act. There are two key points to note in relation to this, firstly, your AI Provider will have set explicit instructions as to how to use the AI system. They will be legally obliged to provide these to you under the act. If you change the use case or adapt their AI system in a way that is outside of this agreement, you then can become the AI Provider. Separately, if you modify a third-party AI system, in a way that substantially affects its performance, you can also take on the responsibilities of 'AI Provider'. What constitutes a substantial modification is not yet defined. However, if the AI system or model outputs are changing significantly based on the data, weighting of feature selection, or the thresholds set by your organization, it is possible that this will enable you to be considered the AI Provider of that system. If your original AI Provider has not explicitly restricted you from adapting the system to a high-risk use case, they will be required to provide you with as much information as reasonably possible to ensure you can comply with your new role as an AI Provider.

Key Risk Factor 2:

Model Training - Continuously Retrained versus Batch Training

Risk management in AI systems must carefully consider the approach to model training, particularly when comparing continuously retrained systems with batch-trained ones. AI systems which are continuously retrained, meaning that they take data being fed into the system on an ongoing basis and update model performance in real-time, have an increased risk associated with them. While these systems have more agility, they introduce unique risks, such as the potential for model drift leading to unpredictable behavior, increased vulnerability to adversarial attacks, and challenges in maintaining compliance with regulatory standards due to the tests of maintaining effective oversight.

Models trained in batches follow a similar development and deployment structure to software or website development. They typically follow the standard protocols in place for organizations that are releasing a new version of a website, or piece of software, which includes regression testing, and the ability to roll back to the previous version if issues are identified during testing or deployment. These processes can mitigate risks associated with negative impacts on model performance, security, and compliance. Effective risk management must balance the need for adaptability in AI systems with oversight mechanisms, ensuring that models remain safe, reliable, and aligned with health, safety, and fundamental rights throughout their lifecycle. Where possible, batch training offers a lower risk than models that are getting real-time training.

Key Risk Factor 3:

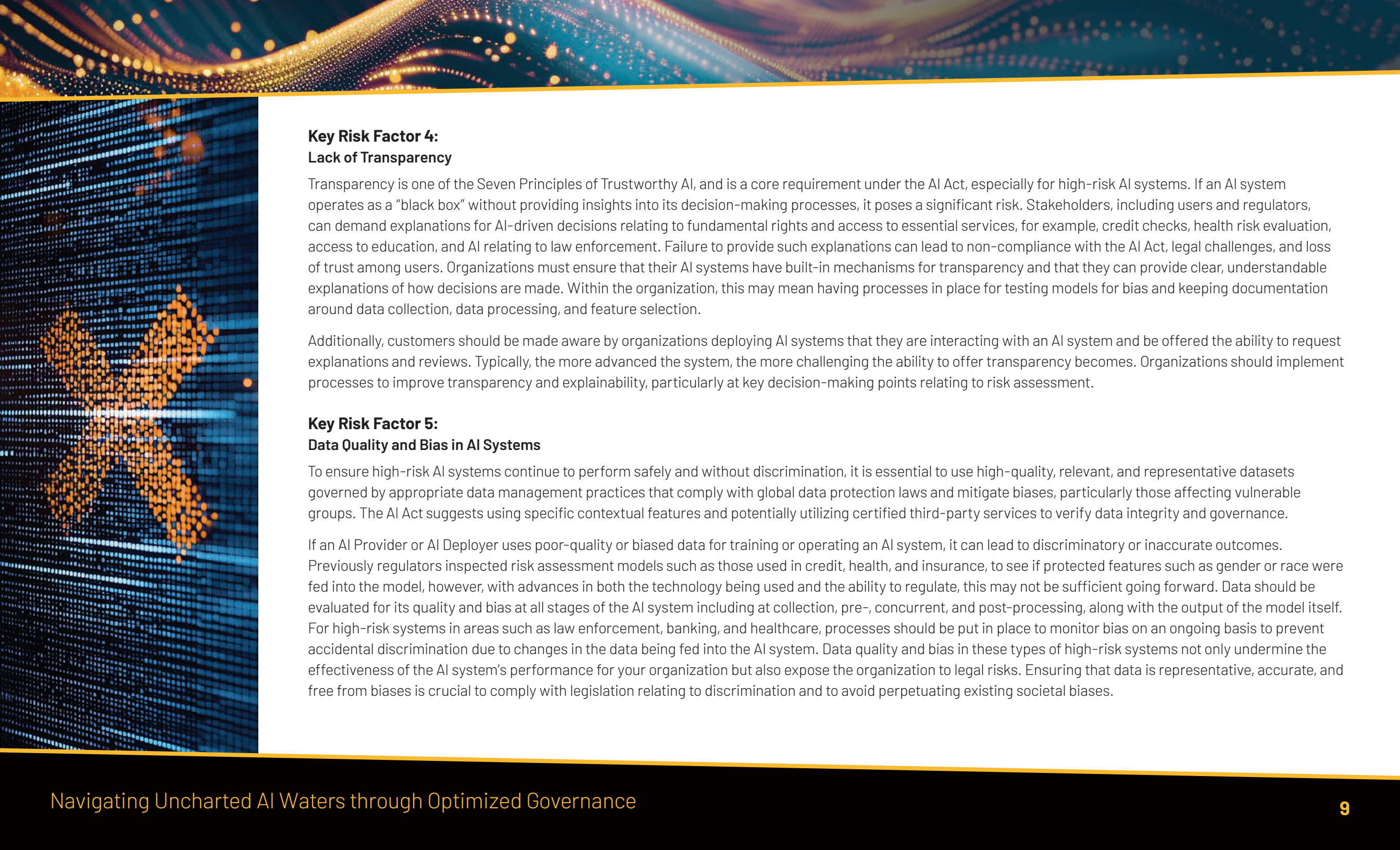
Crossing into Prohibited Practices (Article 5)

Even within a simple machine learning algorithm, there are instances where you could be utilizing AI in a way that crosses over from simple non-compliance into a prohibited practice, meaning the fine springs from 3% to 7% of annual worldwide revenue.

Article 5 (1b) states that prohibited practices include “the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behavior of that person or a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm”.

Due to the level of risk to your organization, you must consider the wording of each point within Article 5 carefully for your use case. For example, if you are deploying AI systems designed to price and advertise high percentage loans, you must look at how the pricing is established, and how the AI-powered advertising system of these loans is finding potential customers. It may be the case that due to how the AI system is trained, it is purposefully designed to target specific disadvantaged groups, for example, single parents on low incomes, or persons with an addiction to gambling, to offer them deceptively high percentage loans, in a way that is likely to cause them significant harm, this could be argued to be a prohibited practice. If you are training models, on large datasets, or using third-party data to identify these disadvantaged groups, in a way that could be argued as manipulative, you must show that you are doing your due diligence, to protect these groups. Breaches of Article 5 are particularly important as it stipulates that both the AI Provider and AI Deployer are responsible if the system is deemed to have crossed into the territory of a prohibited practice.





**Key Risk Factor 4:
Lack of Transparency**

Transparency is one of the Seven Principles of Trustworthy AI, and is a core requirement under the AI Act, especially for high-risk AI systems. If an AI system operates as a “black box” without providing insights into its decision-making processes, it poses a significant risk. Stakeholders, including users and regulators, can demand explanations for AI-driven decisions relating to fundamental rights and access to essential services, for example, credit checks, health risk evaluation, access to education, and AI relating to law enforcement. Failure to provide such explanations can lead to non-compliance with the AI Act, legal challenges, and loss of trust among users. Organizations must ensure that their AI systems have built-in mechanisms for transparency and that they can provide clear, understandable explanations of how decisions are made. Within the organization, this may mean having processes in place for testing models for bias and keeping documentation around data collection, data processing, and feature selection.

Additionally, customers should be made aware by organizations deploying AI systems that they are interacting with an AI system and be offered the ability to request explanations and reviews. Typically, the more advanced the system, the more challenging the ability to offer transparency becomes. Organizations should implement processes to improve transparency and explainability, particularly at key decision-making points relating to risk assessment.

**Key Risk Factor 5:
Data Quality and Bias in AI Systems**

To ensure high-risk AI systems continue to perform safely and without discrimination, it is essential to use high-quality, relevant, and representative datasets governed by appropriate data management practices that comply with global data protection laws and mitigate biases, particularly those affecting vulnerable groups. The AI Act suggests using specific contextual features and potentially utilizing certified third-party services to verify data integrity and governance.

If an AI Provider or AI Deployer uses poor-quality or biased data for training or operating an AI system, it can lead to discriminatory or inaccurate outcomes. Previously regulators inspected risk assessment models such as those used in credit, health, and insurance, to see if protected features such as gender or race were fed into the model, however, with advances in both the technology being used and the ability to regulate, this may not be sufficient going forward. Data should be evaluated for its quality and bias at all stages of the AI system including at collection, pre-, concurrent, and post-processing, along with the output of the model itself. For high-risk systems in areas such as law enforcement, banking, and healthcare, processes should be put in place to monitor bias on an ongoing basis to prevent accidental discrimination due to changes in the data being fed into the AI system. Data quality and bias in these types of high-risk systems not only undermine the effectiveness of the AI system’s performance for your organization but also expose the organization to legal risks. Ensuring that data is representative, accurate, and free from biases is crucial to comply with legislation relating to discrimination and to avoid perpetuating existing societal biases.

Key Risk Factor 6: Insufficient Human Oversight

An AI system that operates autonomously without the possibility of human intervention can lead to harmful outcomes, especially in relation to risk assessment in critical sectors like healthcare, finance, and insurance. The lack of mechanisms for humans to monitor, interpret, and intervene in the AI system's operations increases the risk of non-compliance and potential harm. What this means is that AI Providers should design AI systems with built-in capabilities for effective human oversight, which includes ensuring that humans can override or adjust AI decisions when necessary.

The degree of human intervention varies. AI is typically used to streamline processes, automating manual, laborious tasks. Often, these systems can perform the task in a way that is both more efficient and less biased than if the task was performed by a human. Take the case of risk evaluation for a mortgage. If an organization has initial rules in place, that utilize AI, such as summarization of applications to be handed to a human agent, or instant decision-making by an AI system, they will need to ensure that the automated process has sufficient human oversight. This does not necessarily mean outsourcing every single application for human review, but instead, it means identifying performance metrics, and reviews you need to have in place, to ensure this initial screening is performing as it should and is not biased against individuals or groups. The AI provider is required to supply performance metrics for their systems, you should conduct regular demographic parity tests against your customer base, to ensure that these performance metrics are consistent for protected groups such as gender and race. It is particularly important to remember that the model can infer protected characteristics based on other information provided by the applicant, so even if those characteristics aren't directly entered, you must ensure you are testing for equal outcomes or parity for these groups. In cases where the initial screening can be done effectively without machine learning, instead using a simple if/then rule-based logic, the pressure associated with compliance for AI automation at this stage can be alleviated.

Including human involvement to make decisions, as part of the risk assessment adds a required level of sufficient oversight of these humans. For example, if a person is tasked with manually reviewing and approving AI decisions related to risk assessments, in particular where the AI provides a score relating to the system recommendation, the organization must ensure that the human is implementing proper oversight, and not just agreeing to the AI decisions. If an auditor reviewed approval logs and found a significantly high rate of approval for AI decisions, this would call into question the level of human





PRIVACY, SECURITY & DATA PROTECTION

- Is data collection essential, or can synthetic data be used?
- Do we have sufficient anonymization methods?
- Will this project require updates to existing privacy, security, and data protection practices?
- Will existing processes around consent, opting in/out be sufficient?



FAIRNESS

- Is the data diverse enough for fair representation?
- Will you include processes to detect and correct bias?
- Could specific groups/individuals be disproportionately impacted?
- Could this technology be misused to cause harm to individuals or groups?
- Have you identified and consulted the stakeholders for this system?



TRANSPARENCY

- Will you prioritize model transparency in the design of this system?
- Will affected individuals understand AI-driven decisions?
- Will the model provide explanations for its decisions?



ACCOUNTABILITY & GOVERNANCE

- Who is accountable for the AI's decisions and errors?
- Will you include governance and oversight mechanisms in the design?
- Will you include processes to manage and correct errors?
- Will the system be easily auditable?



HUMAN RIGHTS & AUTONOMY

- Could the system impact individuals' autonomy or decision-making?
- Is there a risk of overreliance on the system and have you taken steps to mitigate it if so?
- Do individuals have control over their data or decisions?
- Could the system cause psychological harm?



ENVIRONMENTAL & SOCIETAL IMPACT

- Does this project consume significant resources?
- Could it reinforce existing social inequalities?
- Who benefits from the system, and who might be harmed?
- Are there long-term societal impacts from this technology?

oversight at this point of the system.

Risk Evaluation & Mitigation

Ensure your organization documents which AI systems you are utilizing. This should include both third-party and in-house systems. It should include everything from simple machine-learning models to advanced LLMs such as AI-driven assistants. Every AI system must be considered for its effect on the individual or specific groups of people, particularly as they relate to safety, health, and fundamental rights. The table below gives an example of the types of AI systems for the use case of an organization operating in the financial services sector. You can consider this example, along with the suggestions provided in the earlier section of this book which looks at categorizing AI systems by risk and utility.

Once systems have been cataloged, roles identified, and appropriate risk classifications identified, you should identify the appropriate mitigations and controls required for these risks. AI technology is complicated and risk mitigation strategies are abundant. The most effective strategy will be to implement a framework such as ethics by design, which incorporates trustworthy AI principles throughout the AI lifecycle. However, as a starting point, consider the following table to help identify and mitigate ethical and safety risks.

DEVELOPING & IMPLEMENTING YOUR AI GOVERNANCE FRAMEWORK

Here's how to develop and implement your AI governance framework by establishing an organizational structure that promotes cross-functional collaboration. It also must set clear objectives aligned with your mission and strategic goals, and extend policies and procedures to encompass AI systems. We emphasize the importance of forming an AI council with representatives from various departments, defining roles and responsibilities, and fostering a culture of ethical AI use to ensure effective management and compliance within your organization.

Establishing an Organizational AI Governance Structure

Many organizations establish an AI council similar to the data protection councils established within organizations between 2016 and 2018 during the rollout period leading up to GDPR. However, unlike GDPR which mandated a role such as a data protection officer (DPO), the AI Act has not insisted on a specific role like this. The AI Act specifically regulates an umbrella term for types of technology that are core to many organizations' ability to operate. We predict that it is likely that the organizational structure within companies will change in a way that we did not see with GDPR, and instead of the arrival of often a single DPO or individual compliance department, there will be a more significant shift in organizational structure. The often-siloed nature of compliance departments will not be efficient for the governance of AI systems, which are already embedded in a cross-functional way in most organizations.


The AI Act requires ongoing performance monitoring of all high-risk AI

systems in line with the EU's post-marketing monitoring template, due for publication by 2nd February 2026. It's likely that these AI councils will not disband after the AI Act comes into force, as was the case with many GDPR councils, but instead, sit elsewhere within the organization. This is important to consider as you establish the objectives of the AI council and think about the evolving role this council may play in the coming years within your organization.

When establishing your AI governance structure, you should ensure there is representation from all stakeholders within your organization and identification and communication with external stakeholders. AI governance is not solely the responsibility of one team such as the IT department, compliance, or data scientists. It requires a collaborative effort across various functions within the organization. Therefore, you should include members from all key departments such as IT, legal, compliance, HR, product, marketing, customer, and operations to provide a holistic perspective.

When establishing a governance structure, it is important to clearly define the roles of each member and what part they play in shaping policy development, risk assessment, compliance monitoring, and decision-making relating to AI systems within your organization. Siloed organizational structures can hinder progress within organizations. You should establish channels for regular communication to ensure that all stakeholders are informed and engaged. Executive sponsorship for AI governance is critical. Leadership should visibly endorse the AI governance framework, allocate necessary resources, and champion a culture of ethical AI use.

Setting Clear Objectives



A key step in developing an AI governance framework is to define clear objectives that align with your organization's overall mission and strategic goals. Developing or deploying AI systems comes with risks, so ensuring this alignment means you can balance AI initiatives with those. When establishing where AI can add the most value, for example, in customer service, risk assessment, or operational efficiency, you should also recognize not just the associated monetary savings, but also the associated risks with that system.

Both existing and future AI systems should have performance metrics determined for them to ensure they operate in a way that is safe and consistent. While the AI Act will require AI Providers to provide performance metrics to be implemented by 2nd August 2026 for most high-risk AI systems, begin establishing these now for AI systems you provide or deploy in your organization. This is both for compliance and to ensure that your organization is using these technologies to the best of its capabilities.

Examples of performance metrics for AI systems range from model accuracy, model and data bias, customer satisfaction and ROI, or return on investment for the organization.

An example of the usage of the types of performance metrics suggested could be a financial institution aiming to enhance its fraud detection capabilities, which might set objectives to reduce fraud-related losses by a certain percentage while ensuring compliance with data protection regulations and anti-discrimination laws and minimizing false positives that could affect their brand through customer dissatisfaction. Performance metrics in line with the goals and compliance obligations of the organization should be established for each scenario.

Establishing/Extending Policies and Procedures

Policies and procedures form the backbone of your AI governance framework. They provide guidelines and standards for AI development, deployment, and management. Your current data management practices should ensure they are fit for that purpose for AI systems. Outline protocols for data collection, storage, processing, and deletion, ensuring compliance with regulations like GDPR. Data collection, processing, and outputs relating to AI systems are fundamentally different than the existing processes for systems that do not utilize AI.

There are additional risks and considerations that are associated with data being utilized, transformed, and created by AI systems that you need to incorporate into existing data management practices. Standards should be created for model development, whether internal or external. These standards should include training, testing, and validation of AI systems, and outline the requirements relating to data lineage and model versions. This is particularly important for high-risk AI systems, which require documentation not just for AI systems overall, but for each individual version of models that are deployed.

Internal governance policies relating to the monitoring of performance and AI systems should also be created to ensure that systems continue to adhere to the required standards. This should include agreement on metrics, risk levels, acceptable thresholds, and the processes that should be put in place for each instance of the performance metrics moving outside of acceptable levels.

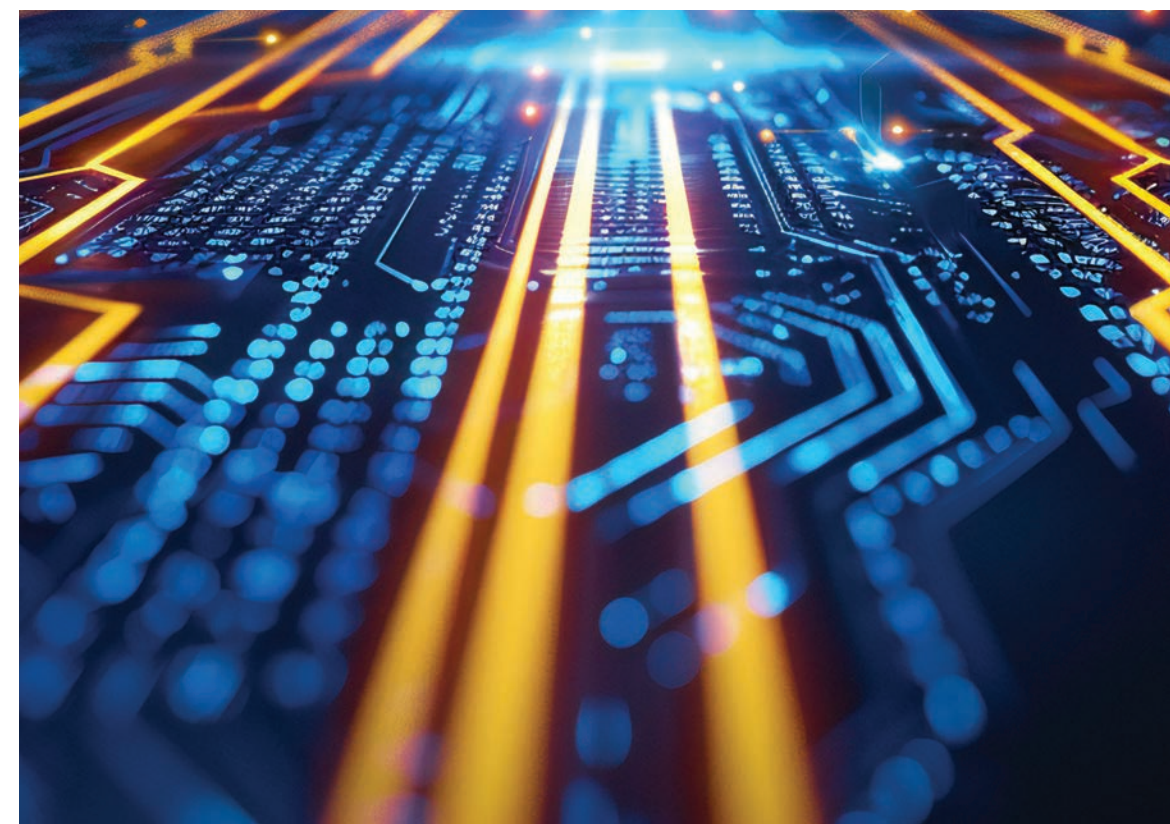
Trustworthy AI guidelines should be established to ensure that AI systems are human-centered, ethical, and safe for usage by society. Implementing practices such as ethics by design can help ensure compliance and that you are developing good AI systems that serve your organization and customers. These guidelines should also ensure that you agree with relevant laws and regulations, such as the EU AI Act. Assign accountability mechanisms to enforce compliance, including regular audits and reporting requirements. Auditability is particularly important for systems that are being changed regularly.

Practical First Steps for AI Governance

- 1. Establish an AI Governance Structure:** Establishing a cross-functional team such as an AI council at this stage is the first step in ensuring compliance for your organization.
- 2. Conduct an AI Inventory:** Compile a list of all AI systems currently in use or planned, including machine learning models, along with their purposes, and data sources. Collecting a high-level table on all AI systems, both internally and externally, should include the following information: AI system owner, development team, the function of the tool (e.g. risk assessment for insurance), the purpose of the tool (e.g. streamline assessments for business efficiencies), direct stakeholders interacting with the tool (i.e. customers, banking agents, AI developers), as well, with a brief description of how it was trained (what data was used, and what model selected) and its associated performance metrics (i.e. accuracy/F1 score, false positives/negatives).
- 3. Perform a Risk Assessment:** Extend the risk assessment table provided in this book to include other risk factors such as potential impact on individuals, data sensitivity, and additional regulatory requirements relating to data, security, and bias.
- 4. Develop a Roadmap:** Create a phased plan for implementing governance measures, and prioritizing high-risk systems within your organization. The key here is to include the development of documentation, including both one-off requirements, and tracking that is to be created on an ongoing basis, identifying performance metrics, thresholds, and AI monitoring solutions you may require, and establishing any larger organizational requirements such as developments to your products, staff training, changes to your organizational structure, or re-evaluation of existing AI vendors.
- 5. Establish Oversight:** For each of your identified systems, establish performance dashboards to track the key metrics identified. Implement processes to flag any issues, including both setting automated security thresholds and mechanisms for feedback loops by those interacting or monitoring the AI system. Where human decision-making occurs within your AI system, you should have processes to ensure that this decision-making is effective and not introducing additional bias into your system. Your organization should also implement processes for regular audits of AI systems to ensure sufficient oversight.

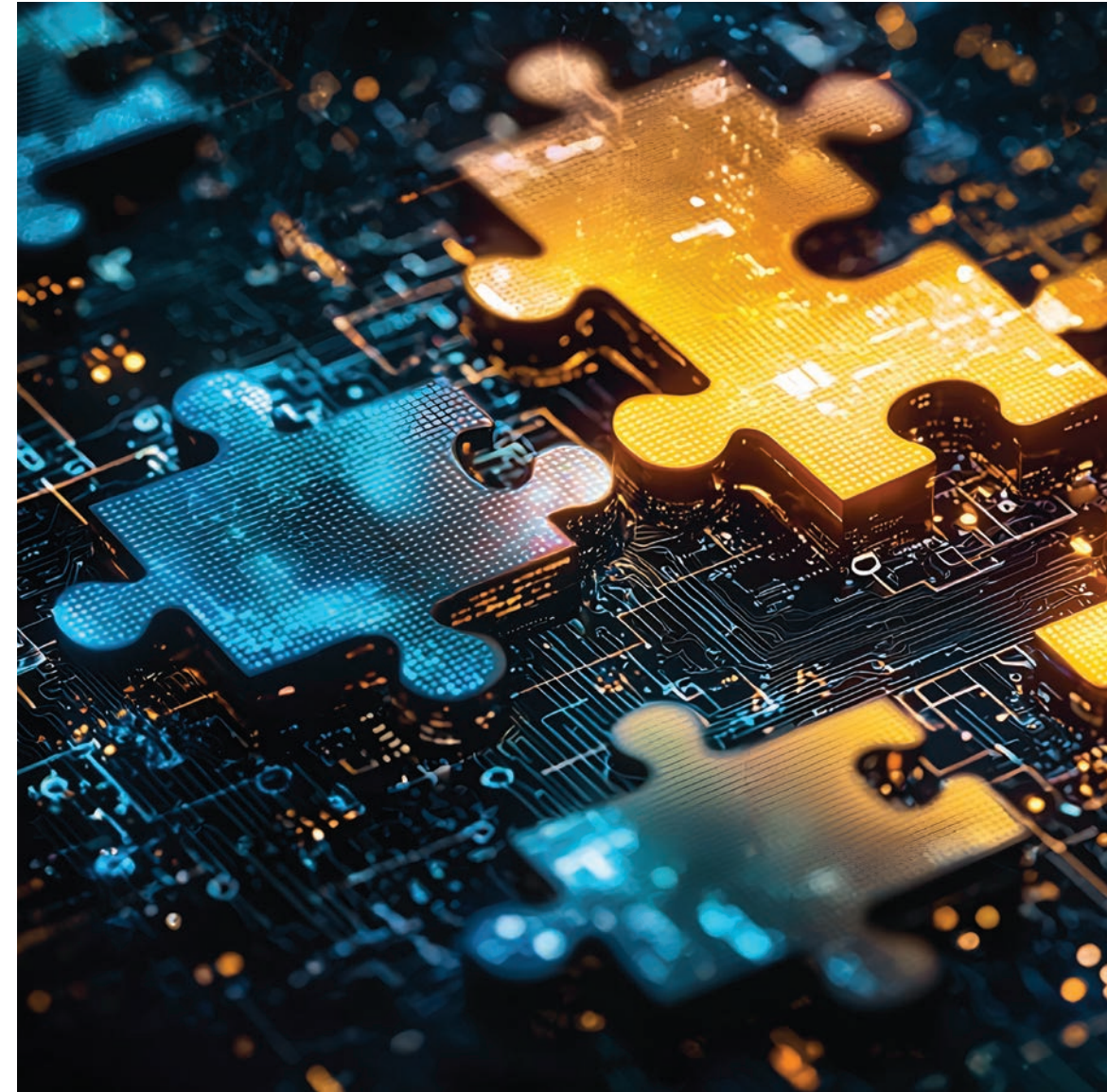
Quick Wins

- **Start with High-Risk Areas:** Focus initial efforts on AI systems classified as high-risk under the EU AI Act or those critical to your business operations.
- **Leverage Existing Frameworks:** Utilize established frameworks and standards like ISO/IEC42001 for AI Management Systems or the AI4People's Institute report on ethics by design to accelerate implementation.



SELECTING THE RIGHT AI PARTNER

Vendors are pivotal in shaping how AI technologies impact your organization. They are responsible for designing, developing, and maintaining AI systems that are ethical, compliant, and effective. Vendors with extensive experience creating or working alongside other high-risk AI systems, particularly those within sectors that are also heavily regulated, can be instrumental in helping your organization navigate the complexities of AI governance. It is vital that the AI Provider and AI Deployer work together to ensure that they are a proper fit to meet best practices in relation to compliance, transparency, ethics, and security is important.



Criteria for Evaluating AI Vendors

Trust is key when choosing an AI vendor that aligns with your organization's needs and values. You may be familiar with the phrase "Trust, but verify"—coined by Russian history scholar, Suzanne Massie, and frequently used by former United States president Ronald Reagan. Organizations want to trust vendors, but they also have a right to expect verification of claims around how security and data processing are being implemented. The reality is that AI development is often siloed, and due to intellectual property and data protection requirements, getting the verification needed to trust AI Vendors can be challenging. However, there are tangible signs that can help to verify some AI vendors' claims and increase trust.

1. Expertise and Experience

Look for vendors with a proven track record in your industry. Their familiarity with industry-specific challenges ensures they can tailor solutions to your context. In high-risk or highly regulated implementations, focusing on vendors with a global presence can bring diverse insights and adaptability to different regulatory environments.

Assess vendor expertise in the AI technologies relevant to your organizational needs, like biometric authentication, machine learning models, risk analysis, or fraud detection algorithms. In areas where AI technology is newly developed such as large language models, it is important to ascertain the risks independently. You should place a particular focus on how you will need to adapt the technology for your requirements and ensure you are maintaining the role of AI Deployer rather than customizing the AI system to the point that your role transitions to AI Provider.

2. Ethical Practices and Transparency

The vendor should integrate ethical and compliance considerations into every stage of AI development, aligning with principles like human agency, fairness, and accountability. If the AI system falls under the high-risk category under the AI Act, the AI Provider should provide documentation, including AI model cards, detailing model design, training data, performance metrics, and potential biases, along with performance metrics, and a risk mitigation plan.

3. Security and Compliance Standards

Ensure the vendor's solutions comply with relevant laws and regulations, such as the EU AI Act and GDPR in Europe, the AI Bill of Rights, and state-specific privacy laws like the CCPA in the United States, and the PIPEDA in Canada. They should adhere to strict data protection protocols, including encryption and secure data handling practices. Look for vendors with certifications like ISO27001, ISO27701, and SOC2, along with those who undergo additional specialist independent audits and certification to verify compliance.

4. Customization and Support

The vendor should offer configurable options to meet your specific requirements, such as adjustable thresholds and performance settings. They should be willing to work closely with your team, providing support for implementation, compliance, and performance monitoring. Evaluate their commitment to customer support, including updates, training, and responsiveness to issues.

5. Innovation and Future-Proofing

Choose vendors that actively enhance their AI models to address new challenges, such as emerging security threats. Their solutions should be scalable to accommodate your organization's growth and evolving AI needs. A vendor investing in research and development, and adherence to not just current, but future compliance requirements demonstrate a commitment to staying at the forefront of AI advancements.

ARTIFICIAL INTELLIGENCE SCREENING QUESTIONS FOR SUPPLIERS: Guard-rails on the Road to AI Assistance

These are meant to help you ascertain the level of knowledge and applicability of that, to your vendor screening quest. Screen, on average three to five selected firms and tabulate their responses to both your business segment and situation.

1 DATA TRANSPARENCY & USAGE

- **What data is/was used to train your models?**
 - Does data include samples that are diverse and representative?
 - Are models bias-tested to ensure fair outcomes across different demographics?
- **How is customer data handled?**
 - Will data be retained or used for training models? (additional questions will apply if so)

2 MODEL AND TECHNOLOGY DETAILS

- **What AI models do you use, and where are they applied?**
 - Inquire about the types of AI models employed and their use cases within the application to ensure they fit Daon's business needs.
- **Do you integrate AI from other vendors?**
 - If so, what checks and balances are in place to prevent these external models from introducing biases into our application?
- **Can we configure how attributes are weighted in your models?**
 - Ensure we can adjust model parameters to align with Daon internal policies and objectives.

3 COMPLIANCE AND GOVERNANCE

- **What is your governance model for data and AI ethics?**
 - Ask for details on how they ensure ethical use of data and AI (includes governance frameworks, ethical guidelines, and any third party audits)?
- **How do you ensure human oversight in critical decisions?**
 - Understand how the system keeps humans in the loop (especially for high-stakes decisions)
 - Confirm if factors can be influenced and weighed in the model's recommendations

4 REGULATORY COMPLIANCE

- **Which AI regulations apply to your product?**
 - Ensure the vendor is compliant with laws in our operational regions, such as the USA, EU, or other applicable jurisdictions.
- **How do you stay updated with global regulatory changes?**

5 PERFORMANCE MONITORING

- **How do you monitor AI model performance, reliability, and bias?**
 - What are the methods for tracking the performance and fairness of their AI models?
- **Can you provide reports and audits from a reputable third party?**

6 SUPPORT AND IMPLEMENTATION

- **What support do you offer for the responsible implementation into Daon systems/environment?**
- **How are issues addressed (e.g. what support channels do you have available)?**

CONCLUSION

As we navigate the uncharted waters of AI innovation, establishing a robust governance framework is both a strategic imperative and a responsibility. By developing organizational structures that promote collaboration, setting clear objectives aligned with your mission, and extending policies to include AI systems, your organization can harness the full potential of AI while mitigating risks. Now is the time to act—begin implementing your AI governance framework at once to ensure ethical, compliant, and effective AI deployment that advances your goals and upholds societal values.



Glossary of Key Terms: EU AI Act

TERM	DEFINITION
EU AI Act	A regulatory framework by the European Union that classifies AI systems based on their risk levels and sets out requirements and obligations to ensure AI technologies are safe, transparent, and respect fundamental rights.
Provider (Role)	A natural or legal person, public authority, agency, or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general purpose AI model developed and places them on the market or puts the system into service under its own name or trademark, whether for payment or free of charge.
Deployer (Role)	Deployer means any natural or legal person, public authority, agency, or other body using an AI system under its authority except where the AI system is used in the course of personal non-professional activity.
Authorised representative (Role)	Any natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general-purpose AI model to, respectively, perform and carry out on its behalf the obligations and procedures established by the AI Act.
Importer (role)	Any natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established outside the Union.
Distributor (role)	Any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market.
Operator (role)	Operators refer to the provider, the product manufacturer, the deployer, the authorised representative, the importer, or the distributor.
High-Risk AI Systems	AI systems classified under the EU AI Act as posing significant risks to health, safety, or fundamental rights, subject to stringent regulatory requirements, including compliance with transparency, data governance, and human oversight provisions.
Prohibited Practices (Article 5)	Under the EU AI Act, certain AI practices are prohibited due to their potential to exploit vulnerabilities or cause significant harm, including manipulation, social scoring, and other activities that infringe on fundamental rights.
Substantial Modification	Significant changes made to an AI system that can alter its performance or intended purpose, potentially shifting the responsibilities under regulations like the EU AI Act from the AI Deployer to the AI Provider if such modifications affect compliance obligations.

Glossary of Key Terms: Trustworthy AI

PRINCIPLE	DEFINITION
1. Human Agency and Oversight	AI systems should support human autonomy and decision-making, ensuring that humans remain in control and can intervene, or override decisions made by AI when necessary. This includes promoting human rights and fundamental freedoms.
2. Technical Robustness and Safety	AI systems should be developed with a focus on reliability, security, and resilience to errors or inconsistencies. They should function correctly under normal circumstances and be able to handle unexpected situations gracefully, minimizing risks of harm.
3. Privacy and Data Governance	AI systems must respect privacy and ensure adequate data protection throughout their lifecycle. This includes secure data handling practices, obtaining proper consent, and ensuring data is used in compliance with relevant regulations like GDPR.
4. Transparency	AI systems should be transparent, providing traceability of their processes and decisions. This includes the ability to explain how and why certain decisions are made, allowing stakeholders to understand and challenge outcomes if necessary.
5. Diversity, Non-discrimination, and Fairness	AI systems should be inclusive, avoiding unfair bias and ensuring that they do not discriminate against individuals or groups. They should be accessible to all and consider the needs of diverse users.
6. Societal and Environmental Well-being	AI systems should benefit all of society, promoting sustainability and ecological responsibility. They should consider their broader impact on society and the environment, contributing positively to societal challenges.
7. Accountability	Mechanisms should be in place to ensure responsibility and accountability for AI systems and their outcomes. This includes auditing and impact assessment processes, as well as the ability to remedy any negative impacts or harms caused by the AI system.

Glossary of Key Terms: Technical

TERM	DEFINITION
AI System	Any software that uses AI techniques such as machine learning, logic-based approaches, or statistical methods to perform tasks that would typically require human intelligence, including perception, reasoning, learning, and decision-making.
Machine Learning (ML)	A subset of AI involving algorithms that improve automatically through experience by using data to learn patterns and make decisions or predictions without being explicitly programmed for each task.
Large Language Models (LLMs)	Advanced AI models trained on vast amounts of text data capable of understanding and generating human-like language, such as OpenAI's ChatGPT, which can produce coherent and contextually relevant text outputs.
Generative Pre-trained Transformers (GPTs)	A type of LLM architecture used in models like ChatGPT, which generates human-like text by predicting the next word in a sequence, based on patterns learned during extensive pre-training on large text datasets.
Black Box Models	AI systems whose internal workings and decision-making processes are not transparent or explainable, making it difficult to understand how inputs are transformed into outputs, posing challenges for accountability and trust.
Model Training	The process of teaching an AI model to make predictions or decisions by exposing it to data and adjusting its parameters to minimize errors, which can be done through batch training or continuous learning methods.
Continuous Learning	An approach where AI models are continuously updated in real-time as new data becomes available, allowing them to adapt to new patterns but also introducing risks of unpredictability and challenges in maintaining oversight.
Batch Training	A method of training AI models in fixed intervals or batches, where the model is updated periodically with new data, allowing for controlled testing and deployment, reducing risks associated with continuous changes.
Performance Metrics	Quantitative measures used to evaluate the effectiveness, accuracy, and reliability of an AI system, including metrics like accuracy, precision, recall, F1 score, and bias measurements, essential for monitoring and improving AI performance.
Data Quality	The condition of datasets used in AI systems being accurate, relevant, complete, timely, without duplicates, representative, and free from biases, ensuring that the AI system's outputs are reliable and do not perpetuate discrimination or inaccuracies.
Bias	Systematic errors in AI outputs resulting from prejudiced assumptions in the AI training process, often due to biased training data, which can lead to unfair or discriminatory outcomes affecting individuals or groups.

Glossary of Key Terms: Governance

TERM	DEFINITION
AI Governance	The framework of policies, procedures, and organizational structures that guide the ethical development, deployment, and management of artificial intelligence systems within an organization, ensuring compliance with regulations and alignment with business objectives and societal expectations.
AI Governance Framework	A structured approach comprising policies, procedures, and organizational structures that guide the responsible development, deployment, and management of AI systems within an organization, ensuring compliance and alignment with ethical standards.
Risk Management	The process of identifying, assessing, and mitigating risks associated with AI systems throughout their lifecycle to ensure they operate safely, ethically, and in compliance with regulatory requirements, especially in relation to health, safety, and fundamental rights.
AI Council	A cross-functional team within an organization responsible for overseeing AI governance, including policy development, risk assessment, compliance monitoring, and strategic decision-making related to AI systems.
Risk Evaluation	The process of assessing the potential risks associated with an AI system, considering factors like impact on individuals, data sensitivity, regulatory requirements, and alignment with ethical standards, to inform mitigation strategies.
AI Vendor Partner	An external entity that provides AI technologies or services to an organization, responsible for ensuring their AI solutions are ethical, compliant, secure, and align with the organization's governance framework and objectives.
Independent Verification	The process of having AI systems evaluated by third-party entities to assess compliance, performance, security, and ethical standards, providing transparency and building trust among stakeholders.
Certifications	Official attestations that an AI system or organization meets specific standards or regulations, such as ISO certifications or compliance with the EU AI Act, demonstrating a commitment to quality, security, and ethical practices.
Ethics by Design	An approach to AI development where ethical considerations are integrated into every stage of the AI system's lifecycle, from design and data collection to deployment and monitoring, ensuring the AI operates responsibly and aligns with societal values.
Security by Design	An approach where security considerations are integrated into every stage of the AI system's development lifecycle, ensuring that the system is robust against threats and vulnerabilities from the outset.
Data Privacy	The protection of personal data from unauthorized access or misuse, ensuring that AI systems handle data in compliance with regulations like GDPR, respecting individuals' rights over their personal information.

Our Top 3 AI Governance Resources

- AI4People Institute’s Report “Towards an Ethics by Design Approach for AI”
<https://ai4people.org/the-new-ai4people-institutes-report/>
- The EU Artificial Intelligence Act
<https://www.europarl.europa.eu/topics/en/article/20230601ST093804/eu-ai-act-first-regulation-on-artificial-intelligence>
- ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system
<https://iso.org/standard/81230.html>