19 practical steps for board members, executives, and IT leaders to implement AI

Written by Rob van der Veer,
Chief AI Officer at Software Improvemenet Group and author of AI standards including ISO/IEC 5338 and the EU AI Act security standard.

4.1 Board AI readiness

An organization's board has a tremendous opportunity to make the organization benefit from AI and is accountable for doing that responsibly. There are plenty of examples of failed AI initiatives that could have been prevented by AI-readiness.

The board must ensure that AI-readiness is understood and implemented across the organization. Let's look at some of the key practices that can help the board succeed.

Step 1: Attain a basic understanding of AI in the board

The board must acquire a basic understanding of AI opportunities, risks, and compliance, as discussed in the previous chapters. This can be achieved, for instance, by integrating AI education into the board's upskilling practice (see further details in later sections). The purpose of this understanding is twofold: to secure management buy-in for AI readiness and to prepare the board for making informed decisions regarding AI.

AI can deliver significant value when applied correctly, but it also presents risks if misused. For example, employees might over-rely on AI, leading to lower quality work that could harm the organization. Imagine a scenario where the sales department uses AI to generate personalized messages for potential customers, only for those messages to contain factual errors.

AI topics for the board to understand:

- Fundamentals of AI technologies: A high-level overview of AI technologies and their capabilities.
- Key legislations: An understanding of the major laws and regulations governing AI, including the responsibilities of board members.
- **AI's impact on individuals, society, and the business: How AI affects various** stakeholders and the potential risks and benefits involved.
- Use cases where AI provides value: Explore how AI is transforming industries and business models through real-world applications.
- AI security basics: The key security considerations associated with AI systems.
- Future AI trends: Insights into upcoming developments and how they might influence the business landscape.

Starting with board education does not necessarily mean other activities such as preliminary AI assessments or exploratory projects must be delayed. However, ensuring that the board has a fundamental understanding of AI early on will make the implementation and integration of AI technologies smoother. It creates a knowledgeable foundation from which detailed and tactical AI strategies can be developed and executed.

Useful resources include:

- A brief introduction to AI and AI readiness (15-minute video)
- Chapter 3 in this document on key legislation
- AI disasters and triumphs, security, privacy, and responsible AI for a broad audience (51-minute video)
- Artificial Intelligence business trends
- 185 real-world Generative AI use cases
- 100+ AI use cases
- AI security overview (13-minute video)
- SANS Institute: AI security made easy (30-minute video)

Step 2: Assign roles and responsibilities

To effectively advance AI initiatives, the first step is to involve key people by clearly defining and assigning roles and responsibilities. This ensures that the practices outlined in this document are properly implemented, leading to the desired results. For example, this includes setting up and executing the AI management system.

- ***Step 2a: Define areas of interest***

  A good way to begin assigning roles is to first define the key areas of interest, aligned with how the organization is currently structured into groups and departments. A starting point would be the areas in which this document's practices are organized: Board, GRC, CISO, and CTO representing the typical departments where these practices would be applied. However, this structure may vary depending on the organization.

  Additional potential areas of interest include:

  - Data management
  - Ethical AI use
  - Regulatory compliance
  - Training and development
  - Procurement and sourcing
  - Innovation management

- *Step 2b: Assign roles and responsibilities*

  Each of the identified areas of interest must have corresponding roles and responsibilities assigned.

  Some of these roles may already exist within the organization, such as:

  - Ethics officer
  - Data governance manager
  - AI security lead
  - Compliance officer
  - Product owner
  - Quality manager
  - Innovation manager

  The practices outlined in this document should be integrated into the responsibilities of these relevant roles.
  A key role is the project or program manager responsible for implementing and maturing AI readiness practices across the organization.
  When assigning responsibilities, consider using the RACI matrix, which identifies:

  - Responsibility (R): Who is responsible for performing the task?
  - Accountability (A): Who is accountable for the outcome?
  - Consulted (C): Who needs to be consulted during the process?
  - Informed (I): Who needs to be kept informed of the progress?

Communicating of roles, responsibilities and contact points

It is important to inform employees on the various roles and to have contact points, such as an 'ai' internal e-mail address or chat channel. This will foster internal alignment, and help to identify needs, impediments and concerns.

AI officer

Organizations may use the title AI officer for various types of roles. Mostly it refers to someone
with leading AI expertise – involved in all or most of the AI initiatives, including AI readiness. Note that strictly speaking, there is no need for a dedicated 'AI officer' role in every organization.

AI responsibilities typically span multiple areas of interest, meaning there doesn't have to be a single person with final responsibility for all AI. This principle helps prevent AI from being treated in isolation and empowers the organization to take initiative. Management of AI initiatives typically happens within the relevant value streams (e.g. product owners shaping AI-related features of a product).

Privacy officers

There is considerable overlap between AI and privacy governance, particularly regarding regulations focused on preventing harm, managing data governance, and conducting impact assessments. Therefore, some AI-related responsibilities may naturally align with the roles of privacy officers or data protection officers.

Step 3: Commit to the principle of building on what is already there

Ensure that all involved parties commit to building upon existing practices (as much as possible), rather than treating AI as a separate initiative. This approach prevents the isolation of AI and avoids the unnecessary proliferation of frameworks. It encourages a holistic approach that leverages what is already in place.

Most organizations already have proven practices in areas such as:

- Risk assessment
- Security awareness training
- Service inventory
- Version control
- Testing
- DevOps
- Knowledge management
- Architecture

To become AI-ready, these existing practices typically require only minor adjustments, making AI governance much less daunting.

When assigning responsibilities, it's important to consider that individuals with an innovation-oriented mindset may be tempted to build new frameworks specifically for AI. However, creating separate AI frameworks is exactly what you want to avoid.

Step 4: Assign a multidisciplinary AI committee

The AI committee should be composed of individuals who have been assigned specific roles and responsibilities or representatives from key areas of interest. Ideally, the committee should include senior leadership representation and expertise in areas such as:

- Research and Development (R&D) (When AI systems are being developed)
- Ethics
- Legal and Compliance
- Privacy
- Security
- Technical AI expertise

If your organization lacks certain expertise, the committee should have access to external resources, such as a legal counsel specializing in AI.

The committee's primary responsibilities include:

- Keeping the organization and stakeholders informed on the status of AI initiatives.
- Holding regular and ad hoc meetings to provide updates, discuss policy decisions, and ensure alignment with organizational goals.

To ensure the committee's responsibilities are clearly defined, include it in the RACI matrix (see previous step). For instance, specify when the AI committee needs to be consulted or informed during AI-related decision-making.

4.2 GRC–Running the AI management system

As discussed earlier, while AI offers significant opportunities, it also introduces great risks and uncertainties that need to be managed. Without proper control, these risks remain invisible, which is why an AI management system is essential. Such a system helps organizations govern AI effectively and avoid potential issues like failed investments, liability claims, fines, reputational damage, and other issues.

There are many examples of issues that could have been prevented with an early evaluation of AI initiatives. One such case is the fraud detection system for Dutch student grants, which faced unavoidable discrimination bias that could have been identified and mitigated. This aspect of AI-readiness is typically assigned to the Governance, Risk, and Compliance (GRC) area of the organization. However, not all organizations have a dedicated GRC group. In such cases, these responsibilities might fall under IT Governance, IT/Internal Audit, IT Internal Control, IT Compliance, the platform team, or risk management.

The primary goal of the AI management system is to ensure that your organization is aware of its AI initiatives, understands the associated risks, and ensures compliance with relevant regulations. The ISO/IEC 42001 standard describes a management system for AI that provides a systematic framework for achieving this. While this standard is not mandatory, it offers valuable guidance on many of the practices involved. Additionally, the AI Exchange provides insights into AI governance.

An AI management system provides a structured approach to overseeing AI and addressing the challenges of AI implementation. Although it may initially seem like a complex administrative task, many organizations already have practices in place that can serve as a foundation, such as risk assessments and inventories of applications. Some of these practices may already be part of the Information Security Management System (ISMS) or integrated into broader corporate GRC frameworks.

Step 5: Identify relevant laws and regulations

Because AI regulations are still evolving, you may believe it's better to wait and see how the regulatory landscape unfolds. However, this couldn't be further from the truth. McKinsey advises organizations to act now to mitigate legal, reputational, organizational, and financial risks associated with AI.

For further details on AI legislation, see Chapter 3 on AI compliance, and consult the OWASP AI Exchange on compliance.

Step 6: Create and maintain an inventory for applications of AI

An inventory of AI applications is crucial for effective AI governance, as it provides a clear overview of the AI-related initiatives within the organization, enabling assessment of both opportunities and risks.

When building this AI inventory, be sure to include:

- *AI innovations*

  These are solutions developed within the organization that rely on an AI model to operate. They can typically be found in portfolio overviews. Make sure to track the following for each AI innovation:

  - The owner: Who is responsible for the AI solution?
  - Applications that use the AI innovation: It's essential to track which applications utilize AI innovations. Establish a policy requiring the documentation of AI models and datasets in systems (e.g., through the SBOM–Software Bill of Materials), even if the AI model was trained in a different initiative or supplied by an external partner.
  - Model cards can be used to standardize how AI models are documented. These cards help ensure proper responses if issues with a specific data source or model are later identified. It's advisable for the GRC team to collaborate with the CTO area on this policy.
  - Involved data sources: What data the AI relies on.
  - Development environment(s): Where the AI model is being developed.
  - Deployments: Where and how the AI model is deployed within the organization.

- *AI innovations plans*

  In addition to current AI innovations, the inventory should also track AI innovation plans, which are ideas or concepts in development (often found in R&D plans). These are potential future applications that may rely on AI models.

- *AI Use in business processes*

  Another critical area to include in the inventory is the use of AI in everyday business processes. For example, the marketing department may be using ChatGPT to generate better website content. These instances can be identified through:

  - Service inventories
  - Requests from employees for permission to use AI tools
  - Surveys sent to teams across the organization to discover existing or planned uses of AI. These surveys are also good opportunities to ask people for concerns, questions, impediments, needs, and ideas – plus they serve as a way to identify individuals for specific AI roles in the organization because of their experience, motivations and affinity.

Implementation guidance

To implement the AI inventory, begin by identifying existing inventories within the organization. Ensure that AI applications are covered in these inventories and label the relevant entries as 'AI', so they can be evaluated during the AI assessment process (detailed in the next section).

Existing inventories that can serve as a foundation include:

- R&D project overview: Current AI-related projects under development.
- R&D roadmap: Ideas and future initiatives that may involve AI.
- Software development portfolio: Ongoing software projects where AI might play a role.
- Inventory of external services: A list of applications and services used by the organization, including those that may involve AI.
- **Asset inventory (as part of risk assessment): This inventory may already include** critical information about AI-related assets.
- Use of external AI services: For example, business processes using toolsl ike ChatGPT. These services might be hard to track, especially if used manually rather than through APIs.

To ensure the AI inventory is comprehensive, it may be helpful to:

- Interview departments to understand their use of AI.
- Send out surveys asking about current or planned AI use, which can also serve to uncover AI development initiatives that may need to be added to the inventory.

The AI inventory can borrow best practices from data governance initiatives, which also focus on identifying data sources, owners, and usage. Similarly, the AI inventory should be regularly evaluated to keep it up to date and ensure it reflects the current AI landscape within the organization.

Step 7: Evaluation of applications of AI

The goal of evaluating AI applications in the inventory is to understand and manage potential consequences for your organization, individuals, and society. This evaluation process serves as a planning and decision-making tool, helping to identify risks and opportunities early.

It is crucial to make it a policy to evaluate AI innovations as early as possible in the idea or design phase. Early evaluation may result in abandoning or significantly changing the idea, which is easier and more efficient to address in the initial stages. Additionally, the evaluation should be conducted regularly as AI initiatives evolve.

For inspiration, see Plot 4 ai for a large repository of AI threats.

*Note: Initially, the number of identified AI uses may seem overwhelming. However, after a proper evaluation (as detailed later in this guide), some uses may be explicitly categorized as "out of scope" due to their low impact. This will allow the organization to focus on high-impact AI applications while minimizing administrative overhead for lower-priority use cases.

Step 7a: Risk assessment

The following types of risks shall be included in risk identification and evaluation while incorporating feedback from relevant stakeholders, and aligning with the organization's values and risk appetite:

- *Assess the risks of harm to individuals and society*
  - Lawfulness and Ethics: Is the AI application lawful and ethical, considering potential harm caused by inaccuracies or manipulation? Consider how the system could impact individuals' health, safety and fundamental rights.
    - o Note that certain regulations (e.g., the EU AI Act) prohibit specific AI applications (see Chapter 3).
    - o While certain risks may not explicitly be regulated, they may still violate human rights or be regarded as morally unacceptable by the organization and/or lead to reputational harm. For example: a company producing deepfake technology for the entertainment industry is faced with abuse of the technology by adversaries and therefore the ethical dilemma of continuing or discontinuing this product.

    Relevant resources and tools to support assessment:
    - Charter of fundamental righs of the European Union
    - ECP AI impact assessment
    - Dutch government AI Impact Assessment
  - Discrimination bias: Is the risk of unfair treatment of protected groups lawful and ethical?
  - Transparency: Can the necessary transparency be provided regarding AI model operations?
  - Right to object: Can individuals be given the right to object to automated decisions?
  - **Data protection:**
    - o Is there a lawful basis for using personal data?
    - o Has consent been obtained where required?
    - o Can individuals request, remove, or update their personal data?
    - o Is data collected for one purpose being reused for another?

*Note: These aspects closely align with privacy governance and may be handled by the Data Protection Officer (DPO). Many of these points are also requirements of privacy legislation, such as the GDPR. Apart from these aspects regarding the impact on individuals, there is also the protection of personal data – which is covered in the CISO section (chapter 4.3).

This impact assessment can help categorize AI applications into risk categories, particularly for regulatory purposes. For example, the EU AI Act identifies several risk categories, each with different rules. Tool to Consider: The Impact Assessment Fundamental Rights and Algorithms, developed by the Dutch government, is a useful tool for evaluating AI applications.

- ***Assess security risks***

  AI systems in use or AI systems that are developed suffer from security risks. AI is software, so all the regular risks apply, plus a number of new risk such as model poisoning, input attacks and specific threats through the supply chain. This may harm individuals and society or the business. Security risks include:

  • Deception risks: Manipulation of the model's behavior in order to deceive through either the input to the model, or through model poisoning which can take place in the supply-chain of model or training data, or during development of an AI system. •

  Disruption risks: Accidental or malicious disruptions that compromise the model's availability.

  • Confidentiality risks: models can be stolen which effective leaks intellectual property, training data typically contains sensitive data and can leak in operation, in development, but also throught the output of the model.

  • For an overview of security threats to the AI system, see the OWASP AI Exchange.

  InGenerative AI systems, a key cyber security risk is leaking of prompts (generative AImodel input), including data that may lead to identification of the user who provided theprompt. The content of the prompt, often in combination with the user can be sensitive information. For this reason, organizations should minimize retention and selection of user data and prompts and perform a risk assessment on how this data is protected by the AI system, whether the system is internal or external. If the system is external (cloud AI), there typically are several options to increase the security level,
  such as opting out of logging and monitoring. It is important to assess whether a cloud AI model is deployed on the virtual private cloud, or on the cloud of the AI vendor. In the latter case, it is key to realise that the sensitive prompt data travels clear text over the internal networkof the AI vendor. In most cases, it will not be used to train the model, but still the risks involved need to be assessed.

  Another typical Generative AI security risk is the confidentiality of 'augmentation data': data that is retrieved to be added to the prompt, as context to specific questions. This is what happens in Retrieval Augmented Generation (RAG) – a very popular architecture forAI systems, where the model is not trained or fine tuned, but documents relevant to a user question are retrieved and then presented to an AI model to help answer the user question. The main security risk is that this augementation data is collected separately and may be out of control of normal data retention and protection measures (e.g. access control), plus: the user may not have the rights to see thisaugmentation data, which is a risk as its content is typically reflected in the output ofthe AI. This can be addressed by applying relevant user privileges when accessing theaugmentation data.

  Formore guidance, see the OWASP AI Exchange risk analysis section.

- ***Assess further risks to the organization***

  While regulations like the EU AI Act focus on individual harm, they may overlook business risks. This is why it is important not to focus only on legislation when evaluating AI, as there may be blind spots that can have significant consequences for the organization. For example, confidentiality breaches or inaccuracies (e.g., false AI-generated meeting notes). Organization risks include:

  - AI inaccuracy consequences - Wrong or low-quality AI system output, typically caused by over-relying on AI, leading to wrong decisions or unhappy customers. Suitable countermeasures are: user education, review/human oversight, automated guardrails, and performance validation tests to detect model drift or model staleness. A form of human oversight is to make sure that AI supports decisions, not make potential harmful decisions autonomously.
  - Price increase of AI services – resulting in altered business cases for AI applications. These increases are likely to happen, as current AI prices are far below even the actual energy cost to run hem.
  - Insufficient expertise to use or develop or identify an AI application can lead to mistakes, failed initiatives or missed opportunities. Evaluate whether the organization has sufficient access to the required knowledge and skills. See also Upskilling in step 10.
  - Stifling policy: policies may prevent users to benefit from AI, for example when it is possible to offer a safe AI system that allows the use of certain sensitive data, and it is not offered.
  - Under-reliance on AI: Avoiding AI and dismissing it too quickly can hamper adoption and prevent the organization from fully realizing AI's potential. This risk is strongly connected to insufficient expertise: users may not be able to use or develop the AI successfully or they may be misinformed about its possibilities. Another reason for under-reliance is the lack of community and sense of support: when there is no co-ordinated effort in learning to use AI in specific workflows and sharing inspirational examples, best practices, and do's and don't's.

- ***Assess the risks to the organization regarding automating human tasks***

  - **Reduced job satisfaction and motivation can be the result of automating certain** tasks with AI.
  - Erosion of skills can be the result of automating certain tasks with AI, which is a problem if those skills are needed to for example review or change the work of AI (e.g. AI used for writing articles, of AI used for generating source code). Stakeholder
  - frustration in the form of clients or others increasingly dealing with AI communication instead of human communication with the organization, either by experiencing reduced quality of communication or information, reduced 'warmth', or by having difficulty with accepting the mere fact. This can lead to reputation damage and loss of business.

Step 7b: Opportunity assessment

- *Assess the business value*

  Evaluate whether the AI application delivers current or potential business value, and if there is a business case in light of the risks and the costs involved. If not, it needs to be assessed if continuation is worth it (e.g. to gain experience with AI), and if embedding it into a value stream would be beneficial.

- *Assess embedding*

  Consider if the AI application is embedded in the organization effectively in terms of who manages it and if there are links with relevant other initiatives. Typically, AI initiatives are managed in the value streams where they add the value.

Step 7c: Act on assessment

- *Decision-making regarding the application*

  Discuss and decide whether the application of AI needs to be discontinued, changed or managed differently in light of the assessment.

  - A way to achieve alignment between initiatives is to have them report to the AI committee or install a separate AI R&D committee to oversee. Such a structure can also be employed for knowledge exchange between initiatives. See also the 'Community of practice' (step 19).
  - Review terms and conditions of involved AI suppliers (if any) regarding potential impact, including indemnifications.

  Let's take the following example. Say that your organization is considering using AI for fraud detection. During the evaluation, it's found that discrimination risks cannot be avoided, making the application illegal. The conclusion would be to discontinue this innovation idea.

- *Propose risk mitigation measures*

  Based on the risks identified in the assessments, propose appropriate mitigation measures. This may involve assessing relevant insurance, based on the identified risks.

- *Update disclaimers, terms & conditions*

  Consider revising the organization's disclaimers, terms, and conditions to protect users (e.g., by informing users of Generative AI systems about how their prompts are handled).

Implementation guidance

This evaluation process may already be part of your organization's risk assessment or privacy impact assessment (PIA/DPIA). In such cases, this practice ensures that the AI-specific aspects are integrated into those existing processes.

Step 8: Communicate the evaluation results and process them further

For each identified use of AI, it is crucial to communicate the evaluation results to the relevant people and departments. Ensure that the results are integrated into policies (see the policy section below) and inform the project teams of their responsibility to address the identified issues. If teams or departments lack the necessary insights or expertise to resolve these issues, provide them with guidance and support.

In some cases, the evaluation may lead to a decision to avoid certain risks, which might mean not pursuing specific goals within an AI application, or even discontinuing the use of AI in certain areas altogether.

Step 9: Create and implement policy

The goal of an AI policy is to manage risks, ensure compliance and maximize AI opportunities while aligning with the organization's goals and values,. This includes providing clear guidance to staff on the approved and prohibited uses of AI tools, along with the conditions for their use. Given the dynamic nature of AI and the evolving laws and regulations, it's essential to review the policy regularly.

Elegant policy implementation

Instead of asking employees to learn and to obey a number of rules from the policy, many of the rules ideally are followed automatically by implementing procedures and tools. For example, if a new policy mandates that training data must always come from a verified source, this can be operationalized by adding a data supplier investigation process to the design approval phase of new machine learning solutions.

Facilitating an AI sandbox

Another example of implementing policy in a facility is to make available a safe and secure AI tool to employees. This will provide a 'sandbox' by protecting the input and output data and by ensuring that an AI model is used with acceptable copyright risks and sufficient accuracy.

This can for example be achieved by implementing an enterprise cloud AI service with opted-out options for the vendor to log and monitor. Such choices depend on the risk appetite. Most organizations will accept AI input (e.g. prompts) being unencrypted in a cloud AI service. Some organizations may opt-in to logging and monitoring of input and output by the vendor. Others will deploy their own AI on-premise, or even prohibit the use of generative AI altogether. By providing a safe environment for employees to experiment, the risk of 'Shadow AI' (users secretly using unapproved AI systems) is mitigated, and grass roots innovation is encouraged.

Input for policy creation

The AI policies should be informed by:

Insights from the evaluation of existing and planned AI applications (as discussed in previous steps).

Industry best practices, such as those in the following resources:

- Generative AI Use at a University
- Responsible Generative AI Use for Research
- Phisher Philips Example Policy for GenAI
- ISACA on Acceptable Generative AI Use Policy

Organization AI principles

Many considerations for policy have to do with norms, values, and strategic views of the organization. This includes typically how to balance certain moral dilemmas. It is recommended to draw these up in a list of AI principles, to be made part of policy documentation and spread in the organization, or even made public, like the AI principles by the HR company Randstad. Related Randstad publications can be found on AI equity and their view on AI and their domain.

Examples of policy considerations

Here are a few examples of questions that can be addressed in policies:

- What classes of data may be sent to cloud AI services?
- For what types of data do we require a self-hosted AI model?
- Should we prohibit certain applications of generative AI due to the risk of hallucinations?
- Is it advisable to use large language models in our public-facing chatbots, given the risk of users provoking offensive language?
- How should we handle potential copyright infringement claims arising from Generative AI output?

Implementation guidance

The most efficient and effective approach is to adapt existing policies (e.g., Information Security Management System (ISMS) policies) by adding specific sections for AI. This could also apply to policies related to third-party applications and services – often part of supplier management. Other examples of existing policies are security requirements and coding guidelines. Depending on the existing policies, a separate AI policy document may not be necessary.

Step 10: Upskilling and establishing a learning organization

To ensure that your organization has the necessary understanding of AI and related policies, particularly among those with AI responsibilities, it can be highly beneficial to implement a knowledge availability program. This program will help make sure there is sufficient understanding in the organization of AI and policy, especially with those that have AI responsibilities assigned.

During such a knowledge availability program you can:

- ***Define AI-related skills and knowledge tailored to the organization:***

  The required knowledge and skills follow from the evaluation (step 7), which is based on the inventory of AI applications (step 6). In addition, part of the policies (step 9) warrant education and instruction of users.

- ***Assess the workforce's current skills:***

  Evaluate the current level of AI knowledge and skills within your workforce. Where necessary, invest in recruitment, or contracting to fill knowledge gaps by involving new people.

- ***Where necessary, develop a tiered education program that provides:***
  - Basic AI literacy for all staff (covering opportunities, AI policy, ethics, legal issues including copyright, policies (step 9), and the risks assessed in AI evaluation – (see step 7), taking current and future AI use into account (step 6). This AI literacy training is not just to educate on how to be careful and compliant – for example by protecting safety, health and individual rights it is also about how to identify great use cases and make AI a success. A great resource is this EU repository of examples of AI literacy programs.
  - Advanced AI training for technical roles (e.g., data scientists, engineers).
  - Consider partnering with accredited institutions to offer certification pathways that validate acquired skills.
  - As much as possible, integrate AI-related content into existing training programs (e.g., security awareness training).

- ***Create a community of practice:***

  Set up a multi-disciplinary community of practice (business strategy, ethics, legal, privacy, security, and technical AI expertise) to learn from each other and share best practices, learnings, methods, or components, on the use of AI and on developing AI systems (see CTO office).

  Organize events like workshops, knowledge-sharing sessions, and datathons (where teams solve data science challenges) to help the organization grow its AI expertise and foster collaboration.

- *Access to external knowledge:*

  Ensure the organization has access to external knowledge and expertise if needed, such as by acquiring legal counsel or partnering with external AI experts.

  Consider collaborations with peers, partners, and other external experts to stay informed about best practices and emerging trends.

Step 11: Stakeholder communication

To effectively communicate with stakeholders, start by defining who they are and creating a communication plan, including regular reports using an agreed-upon format and performance indicators. Engagement is key; collaborating with stakeholders ensures healthy relationships and effective communication, which are crucial for the smooth functioning of your AI initiatives.

Effective communication with stakeholders is essential to the success of your AI initiatives. Begin by identifying key stakeholders and then develop a communication plan that outlines how you will engage with them regularly. The plan should include agreed-upon formats for reporting, as well as key performance indicators (KPIs) to track progress and outcomes.

Stakeholder engagement is critical, as it fosters collaboration, builds trust, and ensures transparency. Regular communication with stakeholders not only helps maintain healthy relationships but also contributes to the smooth functioning of AI initiatives by addressing concerns early and aligning efforts with organizational goals.

Step 12: Implement and improve AI readiness as a program

Run an AI readiness program by implementing the practices and further maturing them. To mature AI readiness, you can develop a clear roadmap that aligns with your AI objectives and policies. This allows you to set clear goals, plan structured improvements, and track progress. The program reports progress to the board and other stakeholders on implementation and
maturation using metrics, e.g. percentage of practices assigned, percentage of practices active, and percentage of roadmap progress. For more information, see the AI Exchange on AI program.

To assess the maturity of your AI-readiness journey and receive recommendations for improvement, you can use a free, quick, and effective assessment tool provided by SIG and EXIN, here.

4.3 CISO taking care of AI security

This chapter covers the AI security practices typically managed by the security office (CISO). AI systems are fundamentally software systems, so their assets, threats, and controls must be included in the organization's broader security management. However, AI introduces a new category of attack, particularly through the input of AI models.

Recent research shows that 77% of businesses have reported a breach involving their AI systems.

A famous example involves ethical hackers successfully fooling an electric car into mistaking stop signs for 35 Miles per hour signs by putting small stickers on the signs.

The relation with privacy

In addition to securing AI systems, the CISO must ensure the information security of personal data, just like it must secure all sensitive data. However, privacy also involves protecting the rights of individuals, which goes beyond technical data security. These rights include:

- Purpose binding: Ensuring data is only used for its intended purpose.
- Consent and lawful basis: Ensuring that the processing of personal data is legally justified.
- Fairness: Making sure AI processes are not biased or discriminatory.
- Right to object: Allowing individuals to object to decisions made by AI systems.
- **Right to remove, update, and transfer personal data: Providing individuals control** over their personal information.

For more details on privacy and security, refer to the Impact Assessment section under GRC in Chapter 4.2, and the OWASP AI Security and Privacy Guide.

Step 13: Incorporate AI assets, security threats, and controls in security

To secure AI systems effectively, it's essential to add AI-specific assets, threats, and controls **to your existing information security management system (ISMS) repositories and policies.**
Note that many of the controls to secure AI are conventional controls (e.g. encryption to protect confidential data) and therefore don't need to be added, as they are already known to the organization.

By incorporating AI-specific elements into your current security management processes, you will drive the extension of risk analysis, awareness, education, security policies, guidelines, tooling, and verification methods for both end users and developers, including AI teams. For further guidance, see the OWASP AI Exchange on AI security management.

We strongly advise against creating or adopting a new, separate security framework for AI.

Instead, focus on building upon your existing security structures to prevent AI from being treated in isolation and to avoid unnecessary duplication. AI-specific assets are:

- Training data
- Test data
- Externally sourced data for training and testing

- The model also referred to as model parameters (values that change when a model has been configured and trained)
- Documentation of models and the development process (including experiments)
- Model input
- Model output (must be treated as untrusted if the training data or model is untrusted)
- Sufficiently correct model behavior, requiring protection of its integrity
  Externally sourced models

Input for policy creation

For a detailed understanding of threats, their impact, and attack surfaces, refer to the OWASP AI Exchange threat frameworks or any of the threat frameworks in the OWASP AI Exchange references.

The OWASP AI Exchange periodic table offers controls that correspond to these threats, which include both process controls and technical controls. Here's how these can be categorized:

- AI is software, so conventional security controls are all relevant.
- Some conventional security controls need slight adaptation, such as rate limiting to prevent malicious experiments and monitoring to detect suspect patterns of use.
- Protection of the development environment including data segregation and being discrete on technical details is essential to protect the integrity and confidentiality of development-time assets, such as training data, to prevent attacks such as data poisoning.
- AI-specific supply chain management is an important control, as the AI supply chain is more complex with train/test data and models also being potentially supplied, and with components running in the engineering environment, instead of just in production.
- Minimization of data is important because of the large data attack surface in AI. This includes the destruction of data after some time and obfuscating data.
- Limiting the effects of unwanted model output is important because the model may have been manipulated, e.g. human oversight, guardrails, and continuous validation. This is a shared responsibility with GRC because it is about more than just security. Even without attacks, AI models can be wrong.
- During development, AI engineers have many controls – mostly based on mathematics - to prevent attacks, e.g. adding noise to training data and federated learning.
- During operation, AI engineers also have controls to filter input and detect attacks as they happen.
- Some security controls are completely new, such as input segregation to mitigate indirect prompt injection risks.

Implementation guidance:

- Add AI assets to the existing asset inventory for security management.
- Include datasets requiring lifecycle management (e.g., training and test data) in the data governance list.
- Externally sourced models and data can be added to an existing overview of supplied components, as they need separate lifecycle management.
- Threats can be categorized and added to an existing risk register, grouped by impact and attack surface (e.g., "model behavior is manipulated through use" or "model behavior is manipulated by breaking into the development environment").
  - This can be the ISMS risk register for security, but sometimes organizations maintain an integrated risk register that includes security risks as well as other risks such as 'Unsafe workplace'
- Extend software product security guidelines by referring to the threats and mentioning the controls including implementation guidelines. These guidelines are typically part of the company's set of policies.
- Extend training programs to cover these new AI security guidelines and threats, ensuring that teams understand how to mitigate them.
- As with traditional software security, it's crucial to "shift left" by building security into AI development from the start. AI model engineering should aim to optimize performance within security and privacy boundaries from the outset. This prevents the need for adding security and privacy measures later, which could require a completely different approach to model development.

Step 14: Incorporate security attacks by AI

Add threats by AI (in contrast to threats to AI) to the risk repository, such as, for example, attackers leveraging AI technologies like deepfakes. These threats should be added to the risk repository and regularly evaluated to determine if existing security controls need to be extended or adapted.

For example, security awareness training may need to include deepfakes, so that staff is aware that the voice of the CEO over the phone can be generated by AI. This is just one of many examples of new security threats to consider.

Step 15: Collaborate with GRC

Work closely with Governance, Risk, and Compliance (GRC) roles (as outlined in the GRC section) to establish robust AI governance frameworks and enhance general AI awareness throughout the organization. This partnership will help in identifying potential risks, implementing appropriate controls, and fostering a culture of accountability.

Step 16: Collaborate with CTO

Work with the people from the CTO office to involve, educate, and instruct AI teams on general (secure) software development best practices, and work with all teams on the AI-specific assets, threats, and controls as mentioned in this CISO section. You could consider mixing AI engineers (e.g. data scientists) and conventional software engineers in teams and analyze software for (security) quality issues as a feedback loop.

4.4: CTO Driving AI development

Chief Technology Officers (CTOs) and Chief Information Officers (CIOs) play a pivotal role in ensuring that organizations fully harness the value of AI. Their leadership is crucial in steering AI initiatives toward achieving tangible business value.

A recent McKinsey article on generative AI adoption reminds us that earlier technological waves, such as the internet and mobile technology, saw many organizations engaging in numerous experiments and pilot projects. However, they often struggled to capture significant business value. The same can happen with AI unless it is guided by effective technical leadership.

For AI readiness, the lessons learned from past technological evolutions are highly relevant. Strong technical leadership can help organizations fully capture AI's potential and integrate it seamlessly into their strategic goals. Let's explore the key practices that CTOs and CIOs can apply.

Step 17: Incorporate AI into the system lifecycle

As previously mentioned, extending and building upon existing practices is more effective than creating a separate process for AI. AI should be treated as software with unique characteristics, as outlined in the new ISO/IEC 5338 standard for AI engineering. Instead of creating a new lifecycle, extend existing software lifecycle practices for AI to leverage established processes, tools, and knowledge. For more guidance, see the AI Exchange on the AI development program.

- Treat AI team members like software engineers and involve them in all regular software engineering processes such as training, platform building, knowledge exchange, threat modeling, and social gatherings.
- Collaborate with the CISO to ensure secure software development processes are applied to AI projects. For more information and guidance, see the AI Exchange on the secure AI development lifecycle.
- Consider mixing data scientists with software engineers in teams to optimize support of these goals.
- Measure quality aspects of software, such as maintainability, test coverage, and security, to provide feedback on how teams are implementing software engineering best practices. Offer actionable recommendations based on these findings, but ensure they consider the unique aspects of AI. Avoid blindly applying conventional improvements–AI requires a tailored approach.

- Apply a different level of scrutiny for every level of technology readiness. Experimental code that is short-lived may be two stars, but code that will at some point be used for a production-ready application requires 4 stars – even if it's not going to be running in production. For more information on TRL see this article on production-ready data science code, this Nature article, and this article on TRL cards.

- Before making important decisions, such as selecting an AI technology stack, it's critical to establish governance and policies. This ensures alignment with organizational principles and avoids violations of pre-established guidelines.

Key AI-specific elements to include in the software lifecycle

- Validate AI models continuously - regarding performance (staleness, model drift), unwanted bias, and regarding compliance and impact (see the Evaluation practice in the GRC section).

- Data governance involve data inventory, data roles and responsibilities, quality management, provenance, and general data policies and standards.

  The data inventory keeps track of data repositories and sources on the following aspects:

  - Lineage (origin and processing path, where it is stored)
  - Business owner and data owner – who is responsible for data quality, compliance, and maintenance
  - Sensitivity classification (as it also takes place in asset management for security)
  - Access permissions
  - Clear definitions of each data element
  - Lifecycle information
  - Intended purpose
  - Relevant policies / allowed usage/ compliance aspects for use/data deletion rules
  - Consumers: which teams, systems, third parties use this data, and when relevant consumer requirements (e.g. interoperability)

- Helpful resources for data governance include: •

  DAMA data management body of knowledge •

  DGI Data Governance Framework

- Keep documentation of the data science work including descriptions and design choices of different experiment stages, and failed experiments. Without proper documentation, even the most valuable software programs are likely to fail because of continuity problems. Undocumented AI initiatives are very hard to transfer to new people, as they have no access to the rationale (why data and configuration are the way they are). When things go wrong or must change, they will need to try everything again, with the risk of taking a very long time or failing. AI engineering typically depends on trial-and-error steps.

- Include AI assets in version management, creating traceability of model versions, training data, and configuration. Similar to traditional software development, versioning your AI system development allows teams to work in distributed and asynchronous environments, manage changes and versions of code and artifacts, troubleshoot issues, and perform proper rollback when necessary.
- Manage third-party relationships, including data and model suppliers, ensuring clarity in allocated responsibilities, and risk management. This includes establishing contractual agreements and providing evidence of conformity where relevant.
- Implement controls to limit the behavior impact of models by including human or automated oversight, assigning minimal privileges to models, ensuring transparency, and conducting continuous validation. For more information or guidance, refer to OWASP's AI exchange.
- Educate AI engineers on software engineering best practices, and software engineers on AI development. Collaborate with the CISO to integrate these training efforts into the organization's broader education and upskilling programs (see the Upskilling practice as described in Chapter 4.1.).

*Note that training AI models requires very specific expertise, which is typically attained by recruiting data scientists through schooling of software engineers in AI, due to the mathematical and data-oriented set of skills and knowled required.

Step 18: Manage AI-supported programming

Effectively managing the use of generative AI to support software development (e.g., code generation) by addressing the following critical factors:

- *Copyright issues of generated code*

  Assess the risk of AI inadvertently generating code that is protected by copyright, especially if the AI was trained on such content. While current rulings do not deem this illegal, the legal landscape is evolving, and future court decisions could change this.

  Some vendors offer indemnification against copyright penalties under specific conditions, but it's essential to understand these terms. For more information, see the OWASP AI Exchange on Copyright.

- ***Nonfunctional quality leading to long-term risks***

  Nonfunctional qualities (such as maintainability, performance, and security) must be closely monitored in AI-generated code. Collaborate with the Chief Information Security Officer (CISO) to develop comprehensive strategies that address these long-term risks.

  Regular audits, code reviews, and security assessments should be conducted to ensure that AI-generated code adheres to the organization's standards.

- ***Lack of code understanding***

  Developers may be tempted to skip fully understanding AI-generated code once it appears to work functionally. This poses risks, such as missing documentation and the inability to effectively maintain or modify the code later.

- ***Risks of review neglect***

  The productivity gains from AI tools can lead to hasty or superficial code reviews, where critical issues may be overlooked. This can be particularly harmful in the long term, as unchecked errors may accumulate.

  Foster a culture of thorough code reviews, ensuring that productivity gains are balanced with high-quality code and security standards. Encourage teams to maintain arigorous review process to catch potential issues early.

- ***Educational aspects***

  Encourage a learning environment where engineers can understand both the benefits and limitations of AI tools. Ensure they continue to develop their programming and problem-solving skills by letting them stay actively involved in the act of coding.

  Engineers will need their coding skills to review and make changes. When they rely too much on AI, they risk atrophying their skills.

Step 19: Organize communities around applications of AI

Setting up a Community of Practice (CoP) for building AI applications can cultivate collaboration, knowledge sharing, and innovation among AI practitioners across the organization. This approach can also be applied to the use of AI, especially when it presents complex challenges for users.

To establish a CoP, start by identifying key stakeholders such as data scientists, engineers, and business leaders who share a common interest in AI. Once identified, define clear goals for the community, including enhancing skills, exchanging best practices, and solving common challenges related to AI development. Regular meetings, workshops, and discussions should be encouraged to keep members engaged and continuously learning from one another. This fosters cross-functional collaboration accelerates problem-solving, and helps standardize AI development practices while ensuring the organization stays at the forefront of advancements in the field.

Alternatively, you might consider establishing a Center of Excellence (CoE). While a CoP is more focused on peer-to-peer learning, a CoE emphasizes standardization and driving innovation through a team of experts. For example, a CoE could spearhead platform engineering for AI by developing reusable components and configurations, potentially justifying the creation of a dedicated AI platform team.