MEST Journal

PROFESSIONAL PAPER
MEST Journal

MANAGING RISKS THROUGHOUT THE ARTIFICIAL INTELLIGENCE LIFECYCLE

Haris Hamidović
Microcredit Foundation EKI and Microcredit Company EKI, Sarajevo, Bosnia and Herzegovina
https://orcid.org/0000-0002-1296-5008

ARTICLE INFO

JEL Category:
M15, D83, L86, O33

Keywords:

artificial intelligence
AI risk
AI lifecycle
risk management
AI security
privacy

ABSTRACT

Artificial intelligence (AI) is increasingly applied in business processes, decision-making, and the delivery of digital services. Unlike traditional information systems, AI systems are highly dependent on data, models, algorithms, and the changing conditions in which they are used. Therefore, AI-related risks cannot be effectively managed only at the time of system development or implementation; instead, organizations need to establish a continuous risk management process throughout the entire AI lifecycle. This paper presents a practical risk management approach structured around the key stages of the AI system lifecycle, including planning and design, data collection and processing, model development and training, testing and validation, deployment, operation and continuous monitoring, change management and retraining, and decommissioning. The proposed lifecycle structure is informed by the OECD AI System Lifecycle approach and adapted in this paper to provide a practical perspective for identifying, assessing, treating, and monitoring AI-related risks. Particular attention is given to data management, security and privacy by design, supplier and third-party risks, model documentation, testing and validation, continuous monitoring, AI asset management, change management, human oversight, and secure decommissioning. The paper aims to support the practical application of lifecycle-based AI risk management, particularly for risk management and information security practitioners who are increasingly required to address AI-related risks.

1 INTRODUCTION

The application of artificial intelligence offers significant opportunities for automation, data analysis, business process optimization, and decision-making support. At the same time, the increased use of AI systems introduces new categories of risk that traditional approaches to information security and IT risk management may not always adequately address. The specific nature of AI systems stems from their dependence on data, models, and algorithms, as well as from the fact that their behavior may change over time. Changes in input data, processing methods,

Address of the author:
Haris Hamidović
haris.hamidovic@eki.ba

Received: 13.08.2026
Revised: 22.08.2026
Accepted: 24.08.2026
Available online: 24.08.2026

models, hyperparameters, or the business context may affect system outputs. Therefore, risk assessment cannot be limited to the development phase but must be treated as a continuous process. AI risk management should therefore be viewed as an integral part of the AI system lifecycle. Risks need to be identified, assessed, and treated before implementation, but they must also be continuously monitored after the system is deployed into production. It is particularly important to define responsibilities, human oversight mechanisms, acceptance criteria, change management procedures, and conditions for retraining or decommissioning the system.

The development of AI-related good practices is influenced by several sources, including academic research, industry standards and frameworks, professional organizations, and legislation and regulatory requirements. These sources provide organizations with principles and practices that can be incorporated into their AI governance and risk management processes. In this context, the use of established frameworks and professional guidance is particularly important because AI regulation is evolving while organizations are already deploying AI systems across a wide range of business processes. Previous work by the author examined AI governance primarily from a governance and regulatory perspective, with particular emphasis on the EU Artificial Intelligence Act and recommendations derived from the ISACA AI governance framework (Hamidović, 2025). That work focused on the broader governance context, regulatory requirements, organizational responsibilities, and the role of established professional frameworks in addressing AI-related risks.

The present paper builds on that earlier work but addresses a different research problem. Rather than focusing primarily on AI governance and regulatory requirements, it examines how AI-related risks can be systematically managed throughout the entire AI system lifecycle. The focus is therefore shifted from the question of what governance and regulatory requirements should be established to the question of how AI risks should be identified, assessed, treated, and monitored at different stages of the AI lifecycle. The paper consequently treats the lifecycle as the organizing structure through which risk management and information security activities can be integrated from initial planning and design through data management, model development, testing and validation, deployment, operation and monitoring, change and retraining, and eventual decommissioning.

This distinction is important because AI risk is not static and may change as the system progresses through its lifecycle. Risks associated with data quality, privacy, security, bias, model performance, human oversight, third-party dependencies, operational use, and changes in the business environment may emerge or change at different stages. Consequently, a risk assessment performed during planning and development may not adequately represent the risk profile of an AI system after deployment. Lifecycle-oriented risk management therefore requires mechanisms for continuous monitoring, reassessment, change management, validation, and treatment of emerging risks.

Accordingly, this paper examines the AI system lifecycle from an integrated risk management and information security perspective, drawing on established good practices and recommendations, particularly those provided by the OECD, NIST, ISO/IEC, and ISACA (OECD, 2024; NIST, 2023; ISO/IEC, 2023; ISACA, 2025). The main contribution of the paper is not the introduction of a new AI development methodology or a new AI governance framework. Rather, it is a practical synthesis that connects AI lifecycle stages with corresponding risk management activities and organizational controls. The proposed approach provides a structured basis for determining what risks should be considered, when they should be considered, and what governance, security, privacy, validation, monitoring, and control activities should be applied throughout the AI system lifecycle.

1.1 Literature Review

The concept of a system lifecycle is well established in software engineering and systems development. Sommerville describes a software process as a set of related activities leading to the production of a software system and identifies four fundamental activities that are present, in different forms, across software processes: software specification, software development, software validation, and software evolution (Sommerville,

2021). In the waterfall model, these activities are represented through distinct phases such as requirements specification, system and software design, implementation and testing, integration and system testing, and operation and maintenance. However, Sommerville also emphasizes that software processes are not universally applicable and that incremental and agile approaches organize these activities differently.

The traditional SDLC therefore provides an important conceptual foundation for understanding the development of AI-enabled software systems. However, AI systems introduce characteristics that require a broader lifecycle perspective. Unlike conventional software, whose behavior is primarily determined by explicitly implemented rules and program logic, AI systems may depend substantially on datasets, trained models, algorithms, hyperparameters, external services, and the operational environment. Changes in these elements may affect system behavior even when the underlying application code has not been modified. The present study therefore treats the AI lifecycle as an extension of the software lifecycle rather than as a completely separate concept.

The distinction between conventional software systems and machine learning-based systems has also been examined from a software engineering perspective. Amershi et al. (2019) identify a number of engineering challenges that arise from the dependence of machine learning systems on data, model development, experimentation, evaluation, deployment, and operational feedback. These characteristics make machine learning systems substantially more complex to engineer and maintain than conventional software systems because changes in data, models, or their surrounding environment can affect system behavior without necessarily requiring corresponding changes to application code. Similarly, Sculley et al. (2015) demonstrate that machine learning systems introduce specific forms of technical debt arising from data dependencies, model complexity, configuration, feedback loops, and changes in the surrounding system environment. These findings support the view that AI lifecycle management should extend the conventional software development lifecycle by explicitly addressing AI-specific dependencies and sources of risk.

The OECD AI System Lifecycle provides an important contemporary foundation for this perspective. The lifecycle approach considers AI systems across interconnected stages rather than restricting attention to development and deployment. The approach adopted in this paper follows the general logic of the OECD lifecycle while adapting it to the practical requirements of AI risk management.

From a risk management perspective, this distinction is important because risk does not necessarily terminate when an AI system enters production. Data quality may change, models may drift, new vulnerabilities may emerge, users may employ the system in ways not originally anticipated, and changes in the business or regulatory environment may alter the risk profile. Consequently, the risk assessment performed during planning and design may differ materially from an assessment performed after deployment. The proposed lifecycle perspective therefore considers risk identification, assessment, treatment, and monitoring as continuous activities rather than as a one-time pre-deployment exercise.

The increasing adoption of AI has resulted in the development of several frameworks and standards addressing AI risk, governance, security, and responsible use. The NIST Artificial Intelligence Risk Management Framework (AI RMF) provides a structured approach for managing risks associated with AI systems, while the AI RMF Playbook provides practical guidance for implementing the framework. (NIST, 2023) These resources are explicitly identified in the ISACA Advanced in AI Risk (AAIR) Review Manual as recommended sources for further study. (ISACA, 2025)

The international standards landscape also provides a foundation for lifecycle-oriented AI risk management. ISO/IEC 23053:2022 (ISO/IEC, 2022) establishes a framework for AI systems using machine learning, ISO/IEC 23894:2023 (ISO/IEC, 2023a) provides guidance on AI-related risk management, and ISO/IEC 42001:2023 (ISO/IEC, 2023b) specifies requirements for an AI management system. These standards provide complementary perspectives covering AI system

architecture, risk management, and organizational governance.

The ISACA AAIR framework further emphasizes that AI risk management should extend beyond technical model development. The recommended practices include consideration of AI risk within contracts and service agreements, supply-chain risk management, ethics, bias, privacy, safety, change management, incident response, business impact analysis, business continuity, disaster recovery, and human oversight at critical decision points.

Similarly, the literature and professional guidance identified in the AAIR Review Manual emphasize accountability and responsible AI. Sources include the IEEE 7000-2021 standard for addressing ethical concerns during system design, ISACA guidance on AI governance, responsible AI, and accountability, as well as COSO guidance on applying its framework to artificial intelligence.

These sources demonstrate that AI risk cannot be reduced to model accuracy or cybersecurity alone. AI risk management requires an integrated consideration of technical, organizational, legal, ethical, privacy, security, operational, and human factors.

A number of AI-specific risks emerge from the dependence of AI systems on data and models. The present paper identifies data management as a foundation of AI risk management and emphasizes data provenance, quality, classification, access control, confidentiality, and protection throughout the data flow. This is consistent with the broader AI risk management literature, which recognizes data quality, bias, privacy, security, and model-related risks as important elements of AI governance.

The importance of lifecycle documentation is also emphasized in the machine learning literature. Gebru et al. (2021) proposed datasheets for datasets as a structured mechanism for documenting the motivation, composition, collection process, recommended uses, and limitations of datasets. Similarly, Mitchell et al. (2019) introduced model cards to provide structured information about the intended use, performance characteristics, limitations, and relevant evaluation conditions of machine learning models. These approaches demonstrate that transparency and risk management should not be limited to the final AI application but should also encompass the datasets and models that constitute critical components of the AI system. From a lifecycle perspective, such documentation provides an evidentiary basis for risk assessment, validation, monitoring, change management, and accountability.

The AAIR Review Manual identifies several areas that extend conventional IT risk management, including AI-related supply-chain risks, contractual requirements concerning data use and intellectual property, human oversight, change management, and the integration of AI risks into incident response and business continuity processes.

Security and privacy also require lifecycle integration. The present paper applies Secure by Design and Privacy by Design principles from the earliest planning and design activities. Privacy considerations are subsequently extended across data collection, model development, training, validation, deployment, monitoring, retraining, modification, and decommissioning.

Another important difference concerns change management. In conventional software systems, change is generally associated with modifications to requirements, source code, configuration, or infrastructure. In AI systems, however, changes to training data, preprocessing procedures, model versions, hyperparameters, or operating context may also modify system behavior. The proposed approach therefore treats significant AI changes as events that may require renewed risk assessment, testing, validation, approval, and documentation.

Continuous learning and retraining further reinforce this distinction. Retraining may introduce changes in system behavior without the conventional notion of a software release. Consequently, retraining should be governed through defined thresholds, validation procedures, approval mechanisms, and documentation rather than being treated as an uncontrolled technical activity.

The reviewed literature, standards, and professional frameworks provide substantial guidance on AI governance, AI risk management, responsible AI, security, privacy, and AI system lifecycles. However, these sources are frequently

presented either as high-level principles and management-system requirements or as specialized technical guidance addressing particular categories of AI risk.

The research gap addressed by this paper is therefore not the absence of AI lifecycle models or AI risk frameworks. Rather, it concerns the practical integration of these perspectives into a single lifecycle-oriented structure that can be used by organizations to identify, assess, treat, and continuously monitor AI-related risks.

The contribution of this paper is a practical adaptation of the OECD AI System Lifecycle from an AI risk management and information security perspective. The proposed approach connects lifecycle stages with concrete risk management activities and organizational controls, including data management, Secure by Design, Privacy by Design, supplier and third-party risk management, model documentation, testing and validation, continuous monitoring, change management, retraining, AI asset management, human oversight, and secure decommissioning. The paper consequently moves from a general description of the AI lifecycle toward an operational risk management perspective that can be used by practitioners.

The proposed contribution should therefore be understood as an integrative and practice-oriented lifecycle framework rather than as a new AI development methodology. Its purpose is to provide organizations with a common structure for determining what risks should be considered, when they should be considered, and what governance or control activities should be associated with different stages of the AI system lifecycle.

1.2 Research Approach

This paper adopts a conceptual and practice-oriented research approach based on a structured review and integration of established AI lifecycle, AI risk management, information security, privacy, governance, and regulatory sources. The objective is not to develop a new AI risk management standard or replace existing frameworks, but to examine how their principles and practices can be integrated into a practical lifecycle-oriented structure for managing AI-related risks.

The analysis draws primarily on the OECD AI System Lifecycle approach, the NIST Artificial Intelligence Risk Management Framework (AI RMF), ISO/IEC 23894:2023, ISO/IEC 42001:2023, and the ISACA Advanced in AI Risk (AAIR) framework and related professional guidance. Relevant regulatory and privacy requirements, including the European Union Artificial Intelligence Act and data protection principles, are also considered where they have direct implications for AI risk management throughout the lifecycle.

The framework development followed four main analytical steps. First, the principal stages of the AI system lifecycle were identified by examining the lifecycle perspectives provided by the OECD and related literature. Second, the principal risk areas associated with each lifecycle stage were identified through comparison of the selected AI risk management standards, frameworks, and professional guidance. Third, relevant risk management activities and governance or control measures were associated with each risk area. Finally, the resulting elements were consolidated into a lifecycle-based AI risk management matrix intended to provide a practical structure for identifying, assessing, treating, and monitoring AI-related risks.

The proposed lifecycle structure consists of eight principal stages: planning and design; data collection and processing; model development and training; testing and validation; deployment; operation and continuous monitoring; change management and retraining; and decommissioning. These stages should not be interpreted as a strictly linear sequence. AI systems may move iteratively between stages, particularly when models are modified, retrained, or subject to significant changes in data, technology, business context, or regulatory requirements.

The analysis also recognizes that certain risk management and governance areas are cross-cutting rather than limited to a single lifecycle stage. These include information security, privacy, human oversight, supplier and third-party management, AI asset management, infrastructure resilience, documentation, and regulatory compliance. Consequently, the proposed matrix associates such areas with the

lifecycle stages in which they are particularly relevant while recognizing that their application may extend across the entire lifecycle.

The resulting framework is therefore intended as an integrative conceptual model rather than an empirically validated predictive model. Its contribution lies in translating principles from established AI risk management and governance sources into a practical structure that helps organizations determine what risks should be considered, when they should be considered, and which governance or control activities may be applied at different stages of the AI system lifecycle. The approach is particularly intended to support practitioners responsible for AI governance, risk management, information security, privacy, and technology management.

2 AI SYSTEM LIFECYCLE AND RISK MANAGEMENT

The AI system lifecycle encompasses the activities and processes involved in the development, deployment, operation, monitoring, and eventual retirement of an AI system. Unlike the traditional Software Development Life Cycle (SDLC), the AI system lifecycle is more iterative and highly dependent on data, models, algorithms, and the context in which the system operates.

The key stages of the AI system lifecycle include:

- planning and design;
- data collection and processing;
- model development and training;
- testing and validation;
- deployment;
- operation and continuous monitoring;
- change management and retraining;
- decommissioning.

This lifecycle-based approach is informed by the OECD AI Principles and the OECD AI System Lifecycle approach, which provide an important conceptual foundation for considering AI systems throughout their lifecycle. The lifecycle structure presented in this paper follows the general logic of the OECD approach but is adapted and further developed from an AI risk management perspective. Figure 1 presents the OECD model, illustrating how the different stages of the AI system lifecycle are interconnected.

From a risk management perspective, it is particularly important to identify and assess risks as early as possible in the AI lifecycle. A risk assessment conducted during the planning and design phase may differ significantly from an assessment performed after deployment, as data, models, users, business processes, and the external environment may change over time. Consequently, AI risk management should be established as a continuous lifecycle process, rather than as a one-time activity performed before implementation.

This means that risks should be continuously identified, assessed, treated, and monitored throughout the lifecycle. Changes to data, models, algorithms, system configurations, business processes, or the operating environment should trigger appropriate risk reassessment and, where necessary, additional testing, validation, retraining, or modification of the AI system. The lifecycle should also include clearly defined criteria for restricting, suspending, or decommissioning an AI system when its risks can no longer be adequately controlled.

The lifecycle approach provides a useful conceptual foundation for understanding where AI-related risks may arise, but practitioners also need practical guidance on how these risks can be identified, assessed, treated, and monitored. (ISACA, 2025)

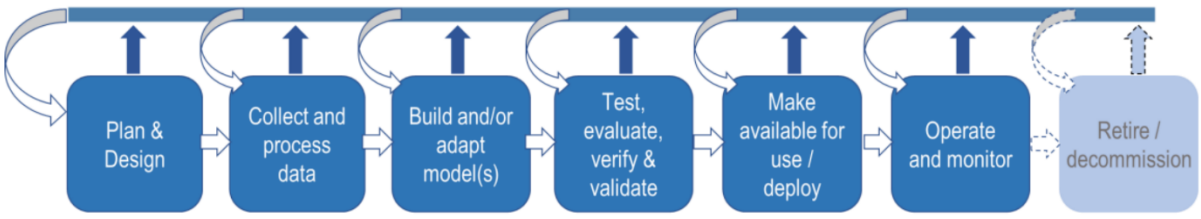

Figure 1. OECD AI System Lifecycle

Source (OECD, 2024)

To operationalize the lifecycle perspective, this paper proposes a conceptual matrix that links each major stage of the AI system lifecycle with the principal risk areas and corresponding risk management activities. The matrix is not intended to replace existing AI risk management frameworks; rather, it provides a practical structure for applying their principles at specific points in the AI lifecycle. In this way, the matrix connects lifecycle management with risk identification, assessment, treatment, and continuous monitoring.

Table 1. *Lifecycle-Based AI Risk Management Matrix*

AI lifecycle stage	Principal risk areas	Key risk management activities	Examples of governance and controls
Planning and design	Business, regulatory, privacy, security, ethical, human oversight, third-party risks	Risk identification and assessment; definition of purpose, ownership, risk criteria and human oversight	AI classification; risk assessment; Secure by Design; Privacy by Design; roles and responsibilities
Data collection and processing	Data quality, bias, privacy, confidentiality, provenance, unauthorized use	Data risk assessment; data classification; provenance and quality assessment; privacy assessment	Access control; encryption; data minimization; retention rules; data lineage
Model development and training	Model performance, bias, security, intellectual property, reproducibility	Model risk assessment; documentation; version control; training-data assessment	Model documentation; versioning; secure development; dataset controls
Testing and validation	Reliability, robustness, bias, security, misuse, regulatory compliance	Independent testing; validation; scenario and edge-case testing; acceptance assessment	Performance metrics; security testing; robustness testing; human oversight
Deployment	Operational, security, privacy, integration and compliance risks	Deployment risk assessment; approval; controlled implementation	Access control; logging; segregation; approval criteria; rollback capability
Operation and continuous monitoring	Model drift, data drift, performance degradation, security incidents, availability	Continuous monitoring; incident detection; reassessment; reporting	Monitoring indicators; alerts; incident response; periodic risk assessment
Change management and retraining	Unintended behavioural changes, new bias, degradation, configuration risks	Change impact assessment; testing; approval; retraining validation	Change records; model/data versioning; approval workflow; rollback
Decommissioning	Residual data, unauthorized access, retained models, dependencies, compliance	Decommissioning assessment; access revocation; retention review; secure disposal	Data destruction; archival; key revocation; removal of APIs and models; audit trail

3 PLANNING AND DESIGN FOR AI RISK MANAGEMENT

During the planning phase, it is essential to clearly define the business problem that the AI system is intended to address, the expected outcomes, the intended users, and the boundaries of its application. The AI solution should be fit for purpose, meaning that it should be designed and developed for a clearly defined, justified, and documented purpose.

At this stage, the organization should also determine whether the use of AI is appropriate for the intended business purpose and identify the

potential consequences of using the system. This includes considering not only technical and operational aspects but also security, privacy, ethical, regulatory, and human-related considerations. Early identification of these factors provides the foundation for effective AI risk management throughout the entire lifecycle.

The initial risk assessment should cover, at a minimum:

- business and operational risks;
- security risks;
- privacy and data protection risks;
- risks related to data quality, availability, integrity, and representativeness;
- bias and discrimination risks;
- ethical risks;
- regulatory and compliance risks;
- risks associated with human decision-making and human oversight;
- risks related to suppliers, third parties, and external AI services; and
- risks associated with the reliability, explainability, and intended use of AI outputs.

Already at this stage, the organization should define the AI system owner, relevant roles and responsibilities, risk acceptance criteria, human oversight requirements, documentation requirements, and criteria for system approval and deployment. Clear ownership and accountability are particularly important because AI systems may involve multiple stakeholders, including business owners, data owners, developers, security and privacy functions, external suppliers, and end users.

Security should be built into the AI system from the outset rather than added after development. The Secure by Design approach requires security considerations to be incorporated into the architecture, data flows, models, applications, infrastructure, and operational processes from the earliest stages of the AI lifecycle.

Security measures should include, as appropriate, protection of data and data pipelines, strong authentication and access controls, encryption, secure system architecture, secure development practices, logging and monitoring, vulnerability management, protection of models and model interfaces, and clearly defined security responsibilities.

AI systems also require consideration of risks that may be specific to their architecture and operation. These may include unauthorized manipulation of training or input data, attacks against models or interfaces, unauthorized access to models or datasets, misuse of AI services, and the generation of unreliable or manipulated outputs. Consequently, security testing and validation should be considered throughout the AI lifecycle rather than being performed only before deployment.

Particular attention should also be given to transparency and auditability. The organization should be able to establish how the AI system was developed, which data and models underpin its operation, which significant changes have been made, and who is responsible for its development, deployment, operation, and oversight. Appropriate documentation, logging, version control, and traceability mechanisms are therefore important elements of AI security and governance.

Privacy should be an integral part of the AI system architecture and should be considered from the earliest stages of planning and design. This approach is consistent with the Privacy by Design principles developed by Ann Cavoukian, which introduced a proactive approach to embedding privacy protection into the design and operation of information systems and technologies more than three decades ago. (Cavoukian, 2009)

The Privacy by Design approach is based on principles such as proactive rather than reactive measures, privacy as the default setting, privacy embedded into system design, full functionality, end-to-end security, visibility and transparency, and respect for user privacy and individual rights. These principles are particularly relevant to AI systems because such systems may involve the collection, processing, combination, and analysis of large volumes of data, including personal data, during model development, training, inference, monitoring, and retraining.

The Privacy by Design approach has become an important element of the modern European data protection framework. It is reflected in the General Data Protection Regulation (GDPR), particularly through the requirements concerning data protection by design and by default. The same approach is relevant to the Bosnia and Herzegovina data protection framework, which

has been developed in alignment with the European data protection approach. Privacy should therefore not be treated solely as a legal compliance requirement but as an integral component of AI system architecture, development, deployment, operation, and governance. (European Union, 2016) (Bosnia and Herzegovina, 2025)

In the context of AI systems, organizations should apply the principles of data minimization, purpose limitation, storage limitation, accuracy, integrity and confidentiality, and accountability. These principles should be considered not only during data collection and processing but throughout the entire AI lifecycle, including model development, training, validation, deployment, monitoring, retraining, modification, and eventual decommissioning.

AI systems may introduce specific privacy risks that should be considered during the planning and design phase. These may include excessive collection of personal data, use of data beyond the original purpose, unauthorized access to training datasets, inappropriate data retention, unintended disclosure of personal information through model outputs, re-identification, inference attacks, and inappropriate secondary use of data.

For this reason, privacy risk assessment should be integrated into the overall AI risk management process from the beginning and revisited whenever significant changes occur to the system, data, model, processing purpose, or operating environment. Where personal data is processed, the organization should also determine whether additional privacy and data protection assessments or safeguards are required under the applicable legal and regulatory framework.

Privacy-Enhancing Technologies (PETs) can provide additional safeguards for reducing privacy risks. Depending on the use case and risk profile, these may include anonymization, pseudonymization, differential privacy, encryption, and other technical and organizational measures designed to reduce the exposure of personal data while maintaining the required functionality of the AI system.

Overall, integrating Secure by Design and Privacy by Design and by Default into the planning and design phase establishes an important foundation for responsible AI risk management. Security and privacy should not be considered as separate controls added after an AI system has been developed. Instead, they should be treated as fundamental design requirements that remain relevant throughout the entire AI system lifecycle.

The lifecycle-based risk management approach is also relevant in the context of the European Union Artificial Intelligence Act (EU AI Act), which establishes a risk-based regulatory framework for artificial intelligence. The AI Act introduces requirements that are closely related to lifecycle risk management, including risk management, data and data governance, technical documentation and record-keeping, transparency, human oversight, accuracy, robustness, cybersecurity, monitoring, and corrective measures for certain categories of AI systems. From a practical risk management perspective, this means that regulatory requirements should not be considered only as a compliance activity performed before deployment. Depending on the role of the organization and the classification of the AI system, relevant requirements may need to be addressed throughout the system lifecycle, from planning and design through development, deployment, operation, monitoring, modification, and eventual decommissioning. The lifecycle approach presented in this paper can therefore serve as a practical structure for mapping applicable regulatory, security, privacy, and governance requirements to specific stages of an AI system's lifecycle. (European Parliament & Council of the European Union, 2024)

4 DATA MANAGEMENT AS THE FOUNDATION OF AI SYSTEMS

The quality of AI systems largely depends on the quality of the data. An organization must understand the origin of the data, the purpose for which it was collected, its quality, classification, sensitivity, and the permitted ways in which it may be used.

AI systems may use different types of data, including:

- numerical data;
- categorical data;
- image data;

- textual data;
- time-series data;
- audio data;
- sensor data;
- video data.

When managing data, particular attention should be paid to the five characteristics of big data: Volume, Velocity, Variety, Veracity, and Value.

The organization should establish mechanisms to ensure the quality, traceability, and provenance of data. It is necessary to document data sources, collection methods, transformations, ownership, and permitted purposes of use.

Particular attention should be given to data containing personal, confidential, financial, or other sensitive information.

Data quality should be assessed based, at a minimum, on the following characteristics:

- completeness;
- uniqueness;
- timeliness;
- validity;
- accuracy;
- consistency.

Issues such as missing values, duplicates, incorrect values, inconsistent formats, and non-representative datasets may lead to reduced model performance or introduce bias.

Data should be protected throughout the entire data flow—from the source system, through data lake/data warehouse environments and data preparation processes, to training, validation, and production use.

Appropriate data classification, access controls, encryption, masking, tokenization, segmentation, and access monitoring should be implemented.

A particular risk arises when data with different classification levels are aggregated in centralized repositories or used in development environments. Therefore, access controls and confidentiality requirements must follow the data throughout the entire AI lifecycle.

5 AI PROCUREMENT AND VENDOR MANAGEMENT

Organizations do not always develop AI solutions internally. They often use commercial AI products, third-party models, API services, or cloud platforms.

The procurement of AI solutions should therefore be integrated into the risk management process.

Contracts should clearly define:

- what data the vendor is permitted to process;
- the purposes for which the data may be used;
- whether the data may be used to train the vendor's models;
- responsibilities for testing and validation;
- security and privacy requirements;
- performance and SLA requirements;
- the right to audit;
- regulatory requirements;
- incident management;
- data and model portability;
- intellectual property rights;
- procedures following contract termination.

Particular attention should be paid to the risk of vendor lock-in, as dependence on a specific model, API, or cloud platform may make migration more difficult and increase operational and financial risks.

6 DEVELOPMENT, TRAINING, AND DOCUMENTATION OF AI MODELS

Model development includes algorithm selection, data preparation, training, hyperparameter selection, and iterative model tuning.

The data should be divided into training, validation, and test sets. Their intended purposes must be clearly defined to enable an unbiased assessment of model performance.

Model Cards represent one approach to the structured documentation of AI models. The documentation should include, at a minimum:

- model name and version;
- owner;
- purpose and intended use;
- limitations;
- architecture;
- datasets used;
- performance metrics;
- known biases;
- security and ethical limitations.

Documentation should be maintained continuously throughout the development process rather than created only at the end of the project.

Centralized documentation management reduces the risk of information loss, particularly when managing a larger number of models and frequent changes. (ISACA, 2025)

7 TESTING AND VALIDATION

Testing and validation must confirm that the AI system meets defined business, technical, security, and regulatory requirements.

Depending on the type of model, metrics such as the following may be monitored:

- accuracy;
- precision;
- recall;
- F1-score;
- robustness;
- latency;
- availability;
- error rate.

Testing should also cover edge cases, unexpected inputs, security scenarios, and misuse scenarios.

For systems that may significantly affect individuals or business decisions, an appropriate level of human oversight should be defined.

8 DEPLOYMENT, MAINTENANCE, AND CONTINUOUS MONITORING

Before production deployment, piloting and controlled testing are recommended. Compatibility with existing infrastructure, applications, and processes should be verified.

The AI system should be continuously monitored after deployment. Monitoring should cover model performance, input data quality, model drift, security events, availability, latency, and other relevant indicators.

Human oversight should be incorporated into critical decision-making points, particularly where an incorrect result could have significant consequences.

9 INFRASTRUCTURE, RESILIENCE, AND SCALABILITY OF AI

AI infrastructure encompasses the computing, software, and network components required for the development, training, and operation of AI systems.

Scaling AI systems increases the volume, variety, and speed of data processing, but at the same time increases risks. Therefore, scaling must not result in reduced security, privacy, or performance.

A modular architecture enables easier scaling and replacement of individual components. Cloud infrastructure provides flexibility but introduces risks such as vendor lock-in, data protection concerns, shared resources, and dependency on service providers.

Secure AI infrastructure should include network segmentation, encryption, key management, redundancy, access control, and appropriate monitoring mechanisms.

10 CHANGE MANAGEMENT IN AI SYSTEMS

Change Management is particularly important for AI systems due to their dependence on data and models.

Changes to data sources, preprocessing processes, models, hyperparameters, or user context may alter system behavior. Therefore, every significant change should be assessed, approved, tested, and documented.

The organization should define:

- criteria for a significant AI change;
- an approval process;
- testing requirements;
- rollback mechanisms;
- documentation requirements;
- responsibilities for emergency changes.

It is particularly important to track the versions of models, datasets, preprocessing processes, and configuration parameters.

11 CONTINUOUS LEARNING AND RETRAINING

Continuous learning enables AI systems to adapt to changing conditions. However, automated adaptation may also introduce new risks, including bias, overfitting, or unintended changes in behavior.

Therefore, retraining should be a controlled process rather than a fully uncontrolled form of automation. For certain systems, thresholds and

conditions that trigger retraining should be defined, together with appropriate validation and approval.

12 AI ASSET MANAGEMENT AND SHADOW AI

AI assets differ from traditional IT assets because a single AI system may include multiple models, model versions, datasets, API services, libraries, and infrastructure components.

The organization should establish an AI system register that includes:

- name and version;
- purpose;
- owner;
- users;
- models and data used;
- vendors;
- licensing and costs;
- regulatory classification;
- security classification;
- lifecycle status.

Shadow AI represents an additional risk because AI tools may be used outside formal governance processes. Identification and management of Shadow AI should include employee awareness and training, periodic assessments, and the establishment of clear acceptable-use rules.

13 DECOMMISSIONING AI SYSTEMS

Decommissioning a system should not be limited to simply disabling the application. It should cover models, data, copies, configurations, keys, API access, connected services, and other related artifacts.

The process should include:

- migration planning;
- revocation of access;
- secure archiving;
- application of retention requirements;
- secure destruction of data;
- removal of models and related artifacts;
- documentation of activities performed.

Where physical deletion is not possible, appropriate cryptographic erasure methods may be applied, while ensuring an adequate audit trail.

14 CONCLUSIONS

Managing risks throughout the artificial intelligence lifecycle requires a continuous and integrated approach rather than a one-time assessment performed before the deployment of an AI system. The analysis presented in this paper shows that AI-related risks may emerge or change throughout the lifecycle as a result of changes in data, models, system configuration, users, business processes, regulatory requirements, and the operating environment. Consequently, risk management should remain an integral part of AI system management from initial planning and design through operation, modification, retraining, and eventual decommissioning.

From a practical perspective, effective AI risk management requires the integration of several complementary areas, including data management, information security, privacy protection, governance, regulatory compliance, human oversight, supplier and third-party risk management, testing and validation, continuous monitoring, and change management. These areas should not be addressed as isolated activities but should be coordinated according to the specific stage and risk profile of the AI system.

The main practical contribution of this paper is the proposed lifecycle-based AI risk management matrix, which connects eight principal stages of the AI lifecycle with relevant risk areas, risk management activities, and examples of governance and controls. The matrix provides organizations with a structured basis for determining what risks should be considered, when they should be considered, and which measures may be applied at different stages of the AI lifecycle. It is intended as a practical complement to existing standards, frameworks, and professional guidance rather than as a replacement for them.

For organizations introducing or expanding the use of AI, the proposed approach emphasizes several practical priorities. These include establishing clear ownership and accountability for AI systems, maintaining an AI asset register, documenting data and models, applying security and privacy principles from the design stage, assessing third-party and supplier risks, conducting appropriate testing and validation, continuously monitoring system performance and

risk indicators, and controlling significant changes and retraining activities. Organizations should also establish clear criteria for restricting, suspending, or decommissioning AI systems when identified risks can no longer be adequately controlled.

Particular attention should be given to the fact that AI systems may change their behaviour without a conventional change to application code. Changes in training data, preprocessing procedures, model versions, hyperparameters, external services, or the operating context may affect system outputs and therefore alter the associated risk profile. Change management, version control, monitoring, and periodic reassessment should consequently be treated as essential elements of AI risk management rather than as optional technical activities.

The lifecycle perspective also provides a practical basis for integrating AI-related risks into existing organizational processes. Rather than establishing completely separate mechanisms for every AI application, organizations can incorporate AI-specific considerations into established processes for information security, privacy, risk management, supplier management, incident response, business continuity, change management, and internal control. Such integration can contribute to a more consistent and sustainable approach to AI governance and risk management.

The proposed approach is conceptual and practice-oriented and should therefore be regarded as a structured practical framework for supporting organizational decision-making rather than as an empirically validated universal model. Its application should be adapted to the type and purpose of the AI system, the organization's risk appetite, the applicable regulatory requirements, and the potential impact of the system on individuals and other stakeholders. Further practical application and case-based evaluation may provide opportunities to refine the proposed matrix and develop more detailed assessment and maturity criteria.

Overall, responsible AI adoption requires organizations to manage not only the risks associated with the initial implementation of AI but also the risks that emerge as AI systems evolve. A lifecycle-based approach enables organizations to maintain continuous visibility, accountability, security, privacy, and control over AI systems and provides a practical foundation for their responsible, secure, reliable, and sustainable use.

REFERENCES

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042.

Bosnia and Herzegovina. (2025). Law on the Protection of Personal Data. Official Gazette of Bosnia and Herzegovina, No. 12/25.

Cavoukian, A. (2009). Privacy by design. Information and Privacy Commissioner of Ontario.

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 – Artificial Intelligence Act. Official Journal of the European Union.

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council – General Data Protection Regulation (GDPR). Official Journal of the European Union, L 119.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Hamidović, H. (2025). Upravljanje umjetnom inteligencijom u kontekstu EU UI regulative i preporuke iz ISACA AAIA okvira. Pravo i finansije, Refam Creative Solutions – REC d.o.o. Sarajevo.

International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). (2022). ISO/IEC 23053:2022 Framework for Artificial Intelligence Systems Using Machine Learning.

International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). (2023a). ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management.

International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). (2023b). ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system.

ISACA. (2025). Advanced in AI Risk (AAIR) Official Review Manual.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1.

OECD. (2024). C/MIN(2024)17 Report on the implementation of the OECD: Recommendation on artificial intelligence. *Meeting of the Council at Ministerial Level, 2-3 May 2024* (pp. 3-41). OECD. Retrieved from https://one.oecd.org/document/C/MIN(2024)17/en/pdf

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28*, 2503–2511.

Sommerville, I. (2021). Software Engineering (10th ed.). Pearson.