AI & Partners

Amsterdam - London - Singapore

EU AI Act

ISO/IEC 23894: 2023

A Practitioner's Roadmap

March 2025

AI & Partners

Sean Musch, AI & Partners

Michael Borrelli, AI & Partners

Charles Kerrigan, CMS UK

Vibhav Mithal, Anand and Anand

AI & Partners

Amsterdam - London - Singapore

AI & Partners defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots professional services, regulatory interventions, and participating in industry groups such as AI Commons, we fight for fundamental rights in the artificial intelligence age.

This report was prepared by Sean Donald John Musch and Michael Charles Borrelli. For more information visit https://www.ai-and-partners.com/.

Contact: Michael Charles Borrelli | Director | m.borrelli@ai-and-partners.com.

This report is an AI & Partners publication.

Contents

Introduction

As artificial intelligence continues to reshape industries, organizations must implement robust risk management frameworks to ensure AI systems are safe, fair, and transparent. ISO/IEC 23894:2023 provides a structured approach to AI risk management, guiding organizations in identifying, assessing, and mitigating AI-related risks throughout the system lifecycle.

This report explores the core principles, implementation strategies, and industry implications of ISO 23894, offering practical guidance for organizations striving to manage AI risks effectively. From ethical considerations to risk assessments, the standard establishes a comprehensive foundation for responsible AI deployment and continuous risk monitoring.

With increasing regulatory oversight—such as the EU AI Act—organizations face growing pressure to demonstrate strong AI risk management practices. By adopting ISO 23894, businesses can enhance stakeholder trust, reduce potential liabilities, and ensure compliance with evolving AI governance requirements.

Whether you are an AI developer, enterprise leader, or policymaker, this report serves as a strategic resource for navigating ISO 23894’s implementation. At AI & Partners, we are dedicated to helping organizations build AI that is ethical, secure, and aligned with international risk management standards.

Best regards,

Sean Musch

Founder/CEO

AI & Partners

Key questions being asked about ISO/IEC 23894: 2023

1. What is ISO 23894?

ISO 23894 is an international standard providing guidance on AI risk management. It helps organizations identify, assess, and mitigate AI-related risks throughout the system lifecycle. The standard emphasizes a structured approach, integrating ethical considerations, transparency, and compliance measures. It applies across industries, ensuring AI is safe, fair, and accountable. By following ISO 23894, organizations can reduce unintended consequences, enhance AI reliability, and meet regulatory requirements. The framework supports continuous risk monitoring and improvement, enabling organizations to proactively address evolving AI challenges while fostering trust among stakeholders.

2. Why is ISO 23894 important for AI risk management?

ISO 23894 establishes a systematic framework for managing AI risks, ensuring responsible AI deployment. AI systems can introduce bias, security vulnerabilities, and ethical concerns, making structured risk management essential. The standard helps organizations develop policies to prevent unintended consequences and ensure compliance with laws. It promotes transparency, fairness, and accountability, addressing societal concerns regarding AI decision-making. Organizations implementing ISO 23894 can improve risk resilience, reduce liability, and enhance trust among users, regulators, and stakeholders. Adopting the standard demonstrates commitment to ethical AI practices and long-term sustainability in AI-driven industries.

3. Who should implement ISO 23894?

ISO 23894 is relevant to any organization developing, deploying, or managing AI systems. This includes technology companies, financial institutions, healthcare providers, government agencies, and manufacturers integrating AI into operations. It benefits organizations seeking to minimize AI-related risks while ensuring compliance with regulations. Businesses using AI for decision-making, automation, or analytics can apply the standard to enhance system reliability and mitigate harm. Regulators and auditors also use ISO 23894 to assess AI governance frameworks. Whether a startup or a multinational corporation, organizations across all sectors can improve AI risk management by adopting ISO 23894.

4. How does ISO 23894 help mitigate AI bias?

ISO 23894 provides guidelines to identify and reduce bias in AI systems. It promotes diverse and representative training data, ensuring AI models do not reinforce discrimination. The standard requires bias assessments throughout AI development, from data collection to algorithm testing. It also encourages transparency, enabling organizations to explain AI decision-making processes. Regular audits and fairness evaluations help detect and mitigate bias in real-world applications. By implementing ISO 23894, organizations can build ethical AI systems that promote fairness, accountability, and trust while complying with anti-discrimination laws and industry standards.

5. What are the key components of ISO 23894?

ISO 23894 consists of multiple components focusing on AI risk identification, assessment, and mitigation. It outlines governance structures, ensuring accountability for AI-related risks. The standard emphasizes transparency, requiring explainable AI decisions and bias detection. Continuous monitoring and improvement mechanisms ensure AI systems remain compliant and ethical over time. Risk treatment plans define actions to address vulnerabilities, while regulatory alignment ensures adherence to legal requirements. ISO 23894 also integrates stakeholder engagement, incorporating diverse perspectives to refine AI risk management strategies and enhance trust in AI applications across industries.

6. How does ISO 23894 support regulatory compliance?

ISO 23894 aligns with global AI regulations, such as the EU AI Act and GDPR, helping organizations meet compliance requirements. It establishes risk management processes that address transparency, fairness, and accountability—key regulatory priorities. The standard mandates regular audits and impact assessments, ensuring AI systems operate within ethical and legal boundaries. By following ISO 23894, organizations can demonstrate responsible AI governance, reducing legal risks and penalties. Implementing the standard also streamlines regulatory reporting, improving readiness for compliance checks and enhancing credibility with customers, investors, and oversight bodies.

7. How can organizations integrate ISO 23894 into AI development?

Organizations can integrate ISO 23894 by embedding risk management principles into AI development workflows. This involves conducting risk assessments during model design, ensuring data integrity, and implementing transparency measures. Organizations should establish governance policies, define accountability structures, and engage stakeholders in risk discussions. Regular audits and AI performance monitoring help maintain compliance. Training employees on ethical AI practices further strengthens implementation. By incorporating ISO 23894 into AI lifecycle management, organizations can enhance reliability, reduce operational risks, and build AI systems that align with ethical, regulatory, and business objectives.

8. How does ISO 23894 address AI security risks?

ISO 23894 provides a structured approach to mitigating AI security risks, such as adversarial attacks, data breaches, and model vulnerabilities. It emphasizes implementing robust cybersecurity measures, including encryption, access controls, and anomaly detection. Regular security assessments help identify weaknesses, ensuring AI systems remain resilient. Organizations are encouraged to develop incident response plans to address potential breaches promptly. By integrating ISO 23894 guidelines, organizations can enhance AI system security, protect sensitive data, and ensure compliance with cybersecurity regulations, reducing the risk of exploitation or manipulation of AI-driven decision-making.

9. What challenges do organizations face in implementing ISO 23894?

Implementing ISO 23894 can be challenging due to resource constraints, lack of AI expertise, and evolving regulatory landscapes. Organizations may struggle with integrating risk management into existing AI workflows. Conducting comprehensive risk assessments and bias evaluations requires specialized tools and expertise. Maintaining continuous compliance amid changing AI technologies also poses difficulties. To overcome these challenges, organizations should invest in AI governance training, leverage automation for risk monitoring, and collaborate with compliance experts. Implementing ISO 23894 gradually through phased adoption can also ease integration and enhance long-term success.

10. What are the benefits of ISO 23894 certification?

Achieving ISO 23894 certification enhances an organization’s credibility, demonstrating commitment to ethical AI practices and risk management. Certification provides a competitive advantage by assuring customers, regulators, and partners of AI system safety and compliance. It helps organizations streamline regulatory processes, reducing legal risks and improving audit readiness. Certified organizations also benefit from stronger AI governance, minimizing operational disruptions caused by unmanaged AI risks. Additionally, ISO 23894 certification fosters trust in AI adoption, enabling businesses to expand AI-driven innovations while maintaining transparency, accountability, and alignment with international standards.

Understanding ISO/IEC 23894:2023

Principles of AI Risk Management

Sub-Clauses:

- N/A

Why?

AI risk management principles help organizations navigate the uncertainties and potential harms associated with AI systems. Without a structured approach, AI can introduce risks such as bias, security vulnerabilities, and unintended consequences. A risk management framework ensures AI is developed and deployed responsibly, minimizing harm to individuals and society. In embedding principles like transparency, lifecycle monitoring, and stakeholder engagement, organizations can enhance trust, compliance, and ethical AI use. These principles provide a foundation for managing AI risks effectively across industries.

What?

AI risk management is a structured approach to identifying, assessing, and mitigating risks throughout an AI system's lifecycle. It involves principles such as a holistic approach—ensuring AI risks are managed within broader organizational risks—and continuous improvement, where AI systems are monitored and refined over time. Transparency and explainability help organizations communicate AI risks and mitigation strategies. Stakeholder engagement ensures diverse perspectives inform risk decisions. These principles work together to ensure AI systems are safe, and ethical

Where?

AI risk management principles apply across all industries using AI, from finance and healthcare to manufacturing and public services. Any organization deploying AI for decision-making, automation, or data analysis benefits from applying these principles. In high-risk sectors like healthcare or autonomous driving, a lifecycle perspective ensures risks are continuously monitored and addressed. Businesses using AI for customer interactions can apply transparency principles to build trust. Ultimately, these principles are relevant wherever AI impacts safety, fairness, security, or compliance.

Who?

AI risk management principles are relevant to organizations of all sizes, as well as regulators, developers, and risk professionals. Business leaders use them to align AI strategies with corporate risk management. Compliance teams apply them to meet regulatory requirements. AI developers incorporate them into model design to ensure safety and fairness. Policymakers and auditors use these principles to assess AI risk frameworks. Whether an AI provider, user, or regulator, all stakeholders benefit from a structured approach to AI risk management.

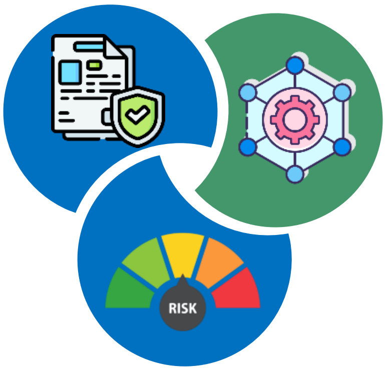
When?

AI risk management principles should be applied throughout an AI system's entire lifecycle, from initial design to deployment, operation, and decommissioning. Risks evolve as AI systems interact with real-world data and environments, requiring continuous monitoring and adaptation. Early in development, risk assessments help identify potential biases or security vulnerabilities. Post-deployment, ongoing evaluations ensure AI systems remain ethical and compliant. AI risk management is also crucial during regulatory reviews, audits, and when AI applications undergo significant modifications or scaling.

How?

AI risk management can be applied by integrating its principles into organizational governance and technical workflows. A holistic approach ensures AI risks are considered alongside broader business risks. Organizations can conduct AI-specific risk assessments, establish governance frameworks, and develop transparency measures to enhance explainability. Engaging stakeholders—such as regulators, customers, and internal teams—ensures diverse perspectives inform risk decisions. Continuous improvement mechanisms, including monitoring tools and regular audits, help organizations adapt to emerging risks and maintain responsible AI practices.

Framework



Sub-Clauses:

- General
- Leadership and commitment
- Integration
- Design
- Implementation
- Evaluation
- Improvement

Why?

AI systems can introduce complex risks, including bias, security vulnerabilities, and unintended societal impacts. Without a structured risk management framework, organizations may struggle to identify, assess, and mitigate these risks effectively. A clear approach to AI risk identification, assessment, treatment, and monitoring ensures AI systems operate safely, ethically, and in compliance with regulations. In applying a risk management framework, organizations can proactively address AI-related uncertainties, build trust with stakeholders, and prevent costly failures or reputational damage.

What?

An AI risk management framework is a structured process for identifying, assessing, mitigating, and continuously monitoring risks associated with AI systems. It includes key components such as risk identification—understanding potential failures, biases, and threats—risk assessment, which evaluates likelihood and impact, and risk treatment, which involves implementing safeguards or modifying AI systems. The framework also emphasizes ongoing monitoring and adaptation, ensuring AI systems remain reliable and compliant as they evolve.

Where?

AI risk management frameworks apply across industries and sectors where AI influences decision-making, automation, or data processing. In healthcare, it helps manage patient safety risks in AI-driven diagnostics. In finance, it mitigates fraud risks in AI-based transaction monitoring. In government, it ensures transparency and fairness in AI-driven public services. Businesses using AI for customer interactions or hiring processes can apply risk assessments to prevent discrimination or bias. Any organization deploying AI benefits from structured risk identification, assessment, and mitigation.

Who?

AI risk management frameworks apply to all stakeholders involved in AI development, deployment, and oversight. AI developers use them to design safer, more reliable models. Compliance and risk management teams apply them to align AI systems with regulatory and ethical standards. Executives and business leaders use them to ensure AI-driven decisions align with organizational risk tolerance. Regulators and auditors rely on these frameworks to assess AI compliance. End users and affected stakeholders benefit from AI systems that are transparent, fair, and secure.

When?

AI risk management should be applied throughout the entire AI lifecycle—from design and development to deployment, monitoring, and eventual decommissioning. Early-stage risk identification helps prevent design flaws before AI models are launched. During deployment, ongoing risk assessments ensure AI operates within acceptable safety and ethical boundaries. As AI systems evolve through updates or new data inputs, continuous monitoring helps identify emerging risks. The framework is also critical when AI applications scale, are repurposed, or face regulatory scrutiny.

How?

Organizations can apply AI risk management by embedding its principles into existing governance structures and technical processes. Risk identification involves analysing AI use cases, data sources, and decision-making logic. Risk assessment applies qualitative and quantitative methods to evaluate potential harm. Risk treatment strategies include modifying AI models, implementing safeguards, or transferring risks through insurance. Ongoing monitoring ensures AI risks are regularly reassessed using key risk indicators. A proactive, structured approach helps organizations manage AI risks effectively and responsibly.

Risk Management Process

Sub-Clauses:

- General
- Communication and consultation
- Scope, context, and criteria
- Risk assessment
- Risk treatment
- Monitoring and review
- Recording and reporting

Why?

AI systems operate in complex environments where risks such as bias, security threats, and unintended consequences can arise. A structured risk management process ensures these risks are systematically identified, assessed, and mitigated. Without it, organizations may face compliance failures, reputational damage, or financial losses. In defining risk scope, assessing vulnerabilities, and applying appropriate treatments, AI risk management enhances transparency, accountability, and trust. It provides a proactive approach to ensuring AI systems remain safe, ethical, and aligned with societal and regulatory expectations.

What?

The AI risk management process is a structured methodology for systematically managing AI-related risks. It involves defining the scope and context of AI risks, identifying potential threats, assessing their likelihood and impact, and implementing risk treatment strategies. This process also includes continuous monitoring, stakeholder consultation, and clear reporting mechanisms. By integrating AI-specific risk criteria, such as transparency, robustness, and societal impact, the process ensures AI systems are developed and deployed responsibly while remaining adaptable to risks.

Where?

The AI risk management process applies across diverse industries, including healthcare, finance, transportation, and government services. It is crucial in high-risk AI applications such as autonomous vehicles, medical diagnostics, and AI-driven decision-making in legal and financial sectors. Organizations using AI for automation, data analysis, or customer interactions can apply risk management to prevent bias, enhance security, and ensure compliance. The process is also relevant for regulatory bodies assessing AI safety and businesses integrating AI into operational risk frameworks.

Who?

The AI risk management process is relevant to organizations of all sizes, including AI developers, risk managers, compliance officers, and executive leadership. Regulators and policymakers use it to evaluate AI system safety and compliance. It applies to engineers designing AI models, legal teams ensuring adherence to AI governance frameworks, and industry auditors assessing AI risks. End users and affected stakeholders also benefit, as effective risk management ensures AI systems operate fairly, securely, and transparently.

When?

AI risk management is a continuous process that applies throughout the AI system lifecycle. It starts at the design phase, where risks are identified and mitigated early. During deployment, ongoing risk assessments ensure AI systems perform reliably under real-world conditions. Continuous monitoring allows organizations to detect emerging risks, adapt controls, and refine AI models over time. The process is also crucial during system upgrades, regulatory audits, or when AI applications are repurposed for new use cases.

How?

Organizations apply AI risk management by integrating it into existing risk governance structures. The process begins with defining AI-specific risk criteria and understanding societal, legal, and operational contexts. Risk assessment involves identifying potential hazards, analysing their impact, and evaluating mitigation strategies. Risk treatment may include modifying AI models, implementing safeguards, or transferring risks through insurance. Ongoing monitoring, stakeholder communication, and clear documentation ensure continuous improvement. A structured and proactive approach enables responsible AI deployment and long-term risk resilience.

Implementing ISO/IEC 23894:2023

Step 1

Establish Governance & AI Risk Framework

Define AI risk management policies, roles, and responsibilities to align with ISO 23894 principles.

☐ Define AI Risk Management Policies and Ethical Guidelines

Develop a Comprehensive AI Risk Policy.

- Identify key AI risk areas, including bias, transparency, and security.
- Draft clear ethical guidelines and governance structures aligned with ISO 23894.
- Ensure policy accessibility and enforceability across all organizational levels.

☐ Assign Roles and Responsibilities for AI Risk Governance

Establish an AI Risk Management Committee.

- Define leadership roles, including AI ethics officers and risk compliance leads.
- Create cross-functional teams to oversee AI governance and accountability.
- Implement a structured reporting process for AI risk escalation and resolution.

☐ Align AI Risk Management with Regulatory Objectives

Integrate AI Risk Governance into Corporate Strategy.

- Map AI risk management goals to business objectives and regulatory mandates.
- Develop a risk assessment framework that aligns with ISO 23894 requirements.
- Ensure executive buy-in and secure resources for ongoing AI governance initiatives.

Step 2

Integrate Risk Management into AI Systems

Embed risk assessment, mitigation strategies, and compliance measures throughout the AI lifecycle.

☐ Implement AI Risk Assessment Methodologies

Develop a Standardized AI Risk Assessment Framework.

- Define AI-specific risk categories, including bias, security, and operational failures.
- Establish risk assessment checkpoints throughout the AI development lifecycle.
- Use quantitative and qualitative methods to evaluate AI system risks.

☐ Establish Monitoring Mechanisms

Implement an AI Risk Monitoring & Incident Response System.

- Deploy real-time monitoring tools to track AI model performance and anomalies.
- Set up automated alerts for compliance violations, security breaches, or system drift.
- Develop an incident response protocol to address AI-related risks proactively.

☐ Integrate Safety Safeguards into AI Model Design

Embed Fairness, Transparency, and Security Measures in AI Development.

- Apply bias detection and fairness validation techniques during model training.
- Ensure AI decision-making processes are explainable and auditable.
- Implement robust security protocols to protect AI systems from adversarial attacks.

Step 3

Conduct Risk Evaluations and Audits

Perform regular assessments, monitor AI performance, and address emerging risks through audits and reviews.

☐ Perform Regular AI System Risk Audits

Establish a structured AI audit program.

- Define audit frequency, scope, and key performance indicators (KPIs).
- Conduct internal and external audits to evaluate AI system compliance with ISO 23894.
- Document audit findings and implement corrective actions for identified risks.

☐ Establish a Reporting System for AI Risk Incidents

Implement an AI Risk Reporting & Transparency Framework.

- Develop standardized reporting templates for AI-related risks and incidents.
- Ensure stakeholders, including regulators and internal teams, have access to risk reports.
- Create a feedback loop to refine AI risk management policies based on reported issues..

☐ Validate AI Model Fairness

Develop a Comprehensive AI Risk Validation Process.

- Conduct stress testing and scenario analysis for AI system performance.
- Utilize adversarial testing to identify vulnerabilities and biases in AI models.
- Compare AI system decisions with human benchmarks to ensure fairness and reliability.

Step 4 Ensure Continuous Improvement & Compliance

Update risk management frameworks, engage stakeholders, and adapt to evolving AI risks and regulations.

☐ Update AI Risk Management Policies

Establish a Periodic AI Policy Review Process.

- Schedule regular policy reviews to incorporate new risks, technologies, and regulations.
- Gather insights from audits, risk assessments, and external regulatory changes.
- Revise AI governance frameworks to address evolving ethical, legal, and operational challenges.

☐ Engage Stakeholders in Ongoing Risk Assessment

Develop a Cross-Functional AI Risk Advisory Committee.

- Include executives, AI developers, compliance teams, and external ethics experts.
- Conduct regular stakeholder meetings to review AI performance and risk trends.
- Integrate stakeholder feedback into AI risk mitigation strategies and policy updates.

☐ Align AI Governance with Legal Standards

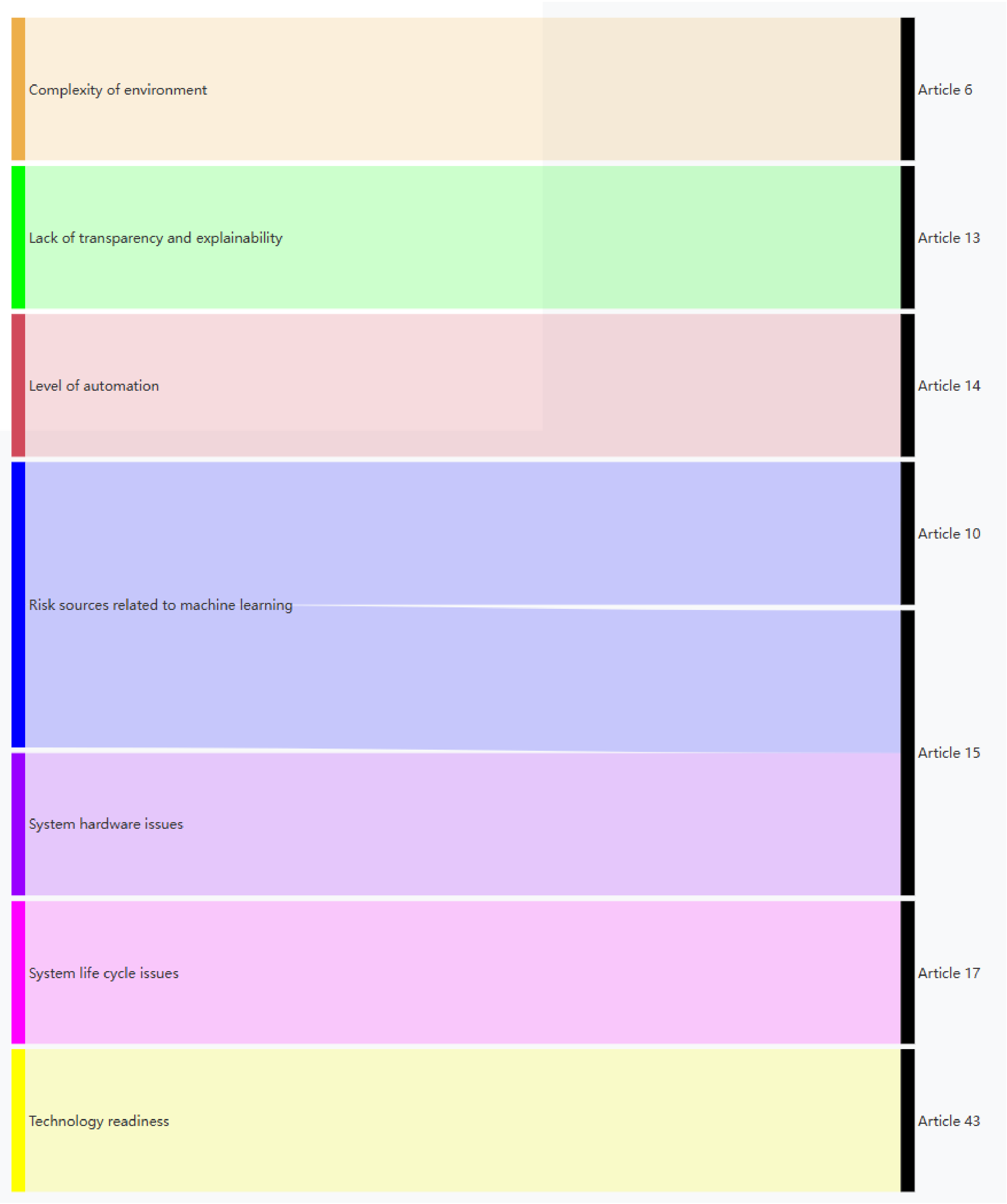
Monitor and Adapt to Regulatory and Industry Developments.

- Track global AI regulations, including ISO updates and industry best practices.
- Conduct gap analyses to ensure AI governance remains compliant with new standards.
- Implement training programs to keep employees informed on AI risk and compliance changes.

Mapping ISO/IEC 23894:2023 to EU AI Act

ISO/IEC 23894:2023

EU AI Act



ISO/IEC 23894:2023			EU AI Act	
Risk Source	**Risk**	**Description**	**Article**	**Explanation**
Complexity of environment	Automated driving	AI-powered driving systems must handle complex road conditions, unpredictable human behaviour, and edge cases, making safety, decision-making reliability, and regulatory compliance critical challenges.	6	The Act classifies AI systems as high-risk if they are used as safety components in products that require third-party conformity assessments, such as automated driving systems. This ensures rigorous evaluation and compliance with safety standards.
Lack of transparency and explainability	Transparency	The ability to make AI operations, decision-making processes, and limitations understandable to stakeholders, ensuring trust, accountability, and regulatory compliance.	13	High-risk AI systems must be designed to ensure transparency, enabling deployers to interpret outputs and use them appropriately. Instructions for use must be clear, complete, and comprehensible.
	Explainability	The property of an AI system that allows humans to understand the key factors influencing its decisions, which is crucial for trust, error analysis, and regulatory alignment.	13	The Act mandates that high-risk AI systems provide information on their characteristics, capabilities, and limitations, including the level of accuracy and robustness, to facilitate understanding by humans.
Level of automation	Human oversight	The involvement of humans in monitoring, intervening, or controlling AI systems to prevent harm, ensure ethical use, and mitigate risks associated with automation.	14	The Act requires high-risk AI systems to be designed with human oversight capabilities, allowing natural persons to monitor and intervene in the system's operation as necessary. This ensures that even fully automated systems can be controlled by humans when needed.
Risk sources related to machine learning	Data quality and collection	Poor data quality, biased datasets, and flawed data collection methods can lead to inaccurate, unfair, or unsafe AI decisions.	10	The Act stipulates that training, validation, and testing data sets must meet quality criteria, be free of errors, and be representative of the intended purpose. It also emphasizes data governance practices to

				manage data collection and processing.
	Continuous learning	AI systems that learn over time may face challenges like data drift, unintended model behaviour, and loss of performance without proper monitoring and safeguards.	15	High-risk AI systems that continue to learn must be developed to minimize biased outputs and ensure feedback loops are addressed with appropriate measures.
System hardware issues	Robustness and cybersecurity	AI systems must be resilient against adversarial attacks, system failures, and security vulnerabilities to maintain reliability and safety.	15	The Act requires high-risk AI systems to be resilient against errors and unauthorized alterations, with technical solutions to ensure cybersecurity and address vulnerabilities.
System life cycle issues	Design and development	Risks arise from inadequate risk assessment, flawed design choices, or insufficient testing during the development phase, impacting AI system performance and safety.	17	Providers must implement a quality management system that includes design, development, verification, and validation procedures.
	Maintenance and updates	Ensuring AI systems remain safe, reliable, and compliant over time requires proper monitoring, patching, retraining, and adaptation to new risks.	17	The Act mandates post-market monitoring and procedures for handling updates and modifications to ensure ongoing compliance and safety.
Technology readiness	Maturity assessment	Evaluating the readiness and reliability of AI technology in a specific application context to ensure it is sufficiently developed for safe and effective deployment.	43	The Act does not explicitly address technology readiness levels but ensures that high-risk AI systems undergo thorough conformity assessments to verify their compliance with safety and performance standards.

Risk Management	AI Risk Management Framework (Art.9)	AI Risk Management Processes (Art.9)				
AI System Lifecycle		Scope, context and criteria	Risk assessment	Risk treatment	Monitoring and review	Recording and reporting
Organisational level activities related to risk management	Governing body sets directions for AI risk management. Top management commits. High-level risk management appetite and general criteria are established.	Feedback reports from AI systems' risk management processes are being received and processed. As a result, the organisational risk management framework is being improved by extending and refining of the organisation'				
		A catalogue of risk criteria.	A catalogue of potential risk sources. A catalogue of techniques for risk sources' assessment and measurement.	A catalogue of known or implemented mitigation measures.	A catalogue of known or implemented techniques for monitoring and controlling AI systems.	A catalogue of established methods and defined formats for tracing, recording, reporting, and sharing the information about AI systems with internal and external stakeholders.
Inception	Governing body examines the AI system objectives in the context of the organisation's and the stakeholders' principles and values. Based on a (typically multi-layer) analysis, determines whether the AI system is feasible and addresses the problem the organisation seeks to solve.	The AI system risk management process and the system's risk criteria are established through customisation of the organisation's risk management framework.	Risk sources specific to the AI system are identified (potentially in a multi-layered manner) and described in detail.	A detailed risk treatment is established. Potentially, "proof of concept", methods are defined.	Necessary "proof of concept" methods are implemented, tested, and evaluated.	The analysis with its results and the recommendation are recorded and communicated to the top management.
Design and development Verification and validation	Governing body continually re-assesses the objectives, the efficacy, and the feasibility of the system based on received feedback reports.	Potentially, the AI system risk criteria is modified as a result of the feedback reports.	The risk assessment is performed continuously (potentially on multiple layers).	The risk treatment plan is implemented. The risk treatment and the (remaining) risks assessment continue until the established risk criteria are met.	During the testing verification, and validation, the risk treatment plan for the system's components as well as for the whole system is assessed and adjusted.	The results are recorded and fed back to the relevant risk management process activities. As necessary, the conclusions are communicated to the management chain and to the governance body.
Deployment	Governing body continually re-assesses the objectives and the feasibility of the system based on received feedback reports.	The AI system risk criteria and the risk management processes are adjusted for the necessary "configuratiom" changes.	The risk assessment is performed continuously (potentially on multiple layers).	The risk treatment plan is potentially updated due to "configuration" changes and implemented. The risk treatment and the (remaining) risks assessment	The AI system's treatment plan is being re-assessed to allow for necessary adjustments.	-

				continue until the established risk criteria are met.		
Operation, monitoring Continuous validation	Governing body continually re-assesses the objectives and the feasibility of the system based on received feedback reports.	Potentially, the AI system risk criteria is modified as a result of the feedback reports.	The system's risk assessment plan is potentially adjusted for risk criteria changes.	The system's risk assessment plan is potentially adjusted for changes in risk assessment outcomes.	The risk treatment plan for the system's components is assessed and adjusted.	
Re-evaluation	Governing body re-examines the AI system objectives and their relation to the organisation's and stakeholders principles and values.	Governing body re-examines the AI system objectives and their relation to the organisation's and stakeholders' principles and values. Based on the analysis, determines whether the AI system is feasible.	The AI system risk management process and the system's risk criteria are re-evaluated against any potential changes to the specific purpose and scope of the AI system, outcome of operation monitoring and new regulatory requirements.	The risk treatment plan is potentially updated. The risk treatment and the (remaining) risks assessments continue until the established risk criteria are met.	The AI system's risk treatment plan is being re-assessed to allow for necessary adjustments.	
Retirement or replacement Triggers a new risk management process with new objectives, risks and their mitigation	Governing body re-examines the AI system objectives based on the analysis, determines whether the AI system retirement or replacement is feasible.	The AI system risk management retirement process and the system's retirement risk criteria are established.	Risk sources specific to the AI system retirement are identified and described in detail.	Detailed risk treatment plan is established.	Necessary "proof of concept" methods are implemented, tested, and evaluated.	

Calls to action

1. Adopt ISO 23894 for AI Risk Management

Ensure your organization is aligned with global AI best practices by implementing ISO 23894, which complements ISO 42001. Strengthen AI governance, manage risks effectively, and enhance compliance with international regulations like the EU AI Act.

2. Conduct an ISO 42001 Pre-Audit

Assess your current AI systems against ISO 42001 with reference to ISO 23894. Identify gaps, mitigate risks, and build a roadmap for compliance to ensure responsible AI deployment.

3. Transparency and Accountability

Implement policies that improve AI system explainability, fairness, and ethical decision-making. Strengthen public and stakeholder trust by embedding transparency and accountability in AI operations.

4. Invest in AI Risk Management and Monitoring

Proactively monitor AI performance, assess impact risks, and integrate continuous improvement processes. Stay ahead of regulatory changes by establishing robust AI risk management frameworks.

5. Partner with AI Governance Experts

Work with AI compliance specialists to navigate ISO 42001 certification with respect to ISO 23894 and regulatory requirements. Leverage expert insights to develop a sustainable, ethical, and compliant AI strategy.

Conclusion

ISO/IEC 23894:2023 represents a significant advancement in structured, ethical, and accountable AI risk management. As organizations worldwide seek to navigate AI-related uncertainties while driving innovation, this standard provides a comprehensive framework for identifying, assessing, and mitigating AI risks throughout the system lifecycle. By establishing clear principles for transparency, accountability, and continuous monitoring, ISO 23894 is already shaping industry best practices and reinforcing global efforts toward responsible AI deployment.

However, effective implementation will determine the true impact of ISO 23894. Organizations face diverse challenges, including aligning AI risk management with existing governance frameworks, ensuring adequate oversight, and balancing regulatory compliance with operational flexibility. Small and medium enterprises (SMEs), in particular, may require additional guidance and resources to integrate ISO 23894 into their AI strategies without compromising competitiveness in an evolving regulatory landscape.

Despite these challenges, early adopters of ISO 23894 are demonstrating the value of structured AI risk management. Leading technology firms, financial institutions, and healthcare organizations are leveraging this standard to enhance compliance, mitigate AI-related risks, and foster stakeholder trust. By embedding ethical safeguards, risk assessments, and lifecycle monitoring into their AI ecosystems, these organizations illustrate how a proactive risk management approach can strengthen both regulatory alignment and operational resilience.

For businesses and policymakers alike, ISO 23894 presents a unique opportunity to lead in AI risk governance. Implementing robust policies, investing in AI risk management capabilities, and fostering cross-sector collaboration will be critical in driving widespread adoption. As industries increasingly rely on AI-driven decision-making, ISO 23894 offers a vital foundation for ensuring AI systems remain fair, secure, and aligned with societal expectations.

Looking ahead, the long-term success of ISO 23894 will depend on industry-wide engagement, ongoing refinement of best practices, and integration with emerging AI regulations, such as the EU AI Act. Organizations that embrace this standard will position themselves at the forefront of responsible AI development, setting a global benchmark for ethical, effective, and sustainable AI risk management.

RISK

About AI & Partners

AI & Partners

Amsterdam - London - Singapore

AI & Partners – 'AI That You Can Trust'

At AI & Partners, we're here to help you navigate the complexities of the EU AI Act, so you can focus on what matters—using AI to grow your business. We specialize in guiding companies through compliance with tailored solutions that fit your needs. Why us? Because we combine deep AI expertise with practical, actionable strategies to ensure you stay compliant and responsible, without losing sight of your goals. With our support, you get AI you can trust—safe, accountable, and aligned with the law.

To find out how we can help you, email contact@ai-and-partners.com or visit https://www.ai-and-partners.com.

Contacts

Sean Donald John Musch, CEO/Founder, s.musch@ai-and-partners.com

Michael Charles Borrelli, Director, m.borrelli@ai-and-partners.com

Authors

Sean Donald John Musch, CEO/Founder

Michael Charles Borrelli, Director

References

European Parliament and The Council of the European Union, (2024), 2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of The Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689 (last accessed 2nd March 2025)

International Organization for Standardization, (2023), 'ISO/IEC 23894:2023: Information technology — Artificial intelligence — Guidance on risk management', accessible at: https://www.iso.org/standard/77304.html (last accessed 3rd March 2025)

International Organization for Standardization, (2023), 'ISO/IEC 42001:2023: Information technology — Artificial intelligence — Management system', accessible at: https://www.iso.org/standard/81230.html (last accessed 2nd March 2025)

Important notice

This document has been prepared by AI & Partners B.V. for the sole purpose of enabling the parties to whom it is addressed to evaluate the capabilities of AI & Partners B.V. to supply the proposed services.

Other than as stated below, this document and its contents are confidential and prepared solely for your information, and may not be reproduced, redistributed or passed on to any other person in whole or in part. If this document contains details of an arrangement that could result in a tax or National Insurance saving, no such conditions of confidentiality apply to the details of that arrangement (for example, for the purpose of discussion with tax authorities). No other party is entitled to rely on this document for any purpose whatsoever and we accept no liability to any other party who is shown or obtains access to this document.

This document is not an offer and is not intended to be contractually binding. Should this proposal be acceptable to you, and following the conclusion of our internal acceptance procedures, we would be pleased to discuss terms and conditions with you prior to our appointment. Images used throughout the document have either been produced in-house or sourced from publicly available sources (see **References** for details).

AI & Partners B.V. is the Dutch headquarters of AI & Partners, a global professional services firm. Please see https://www.ai-and-partners.com/ to learn more about us.

Designed and produced by AI & Partners B.V