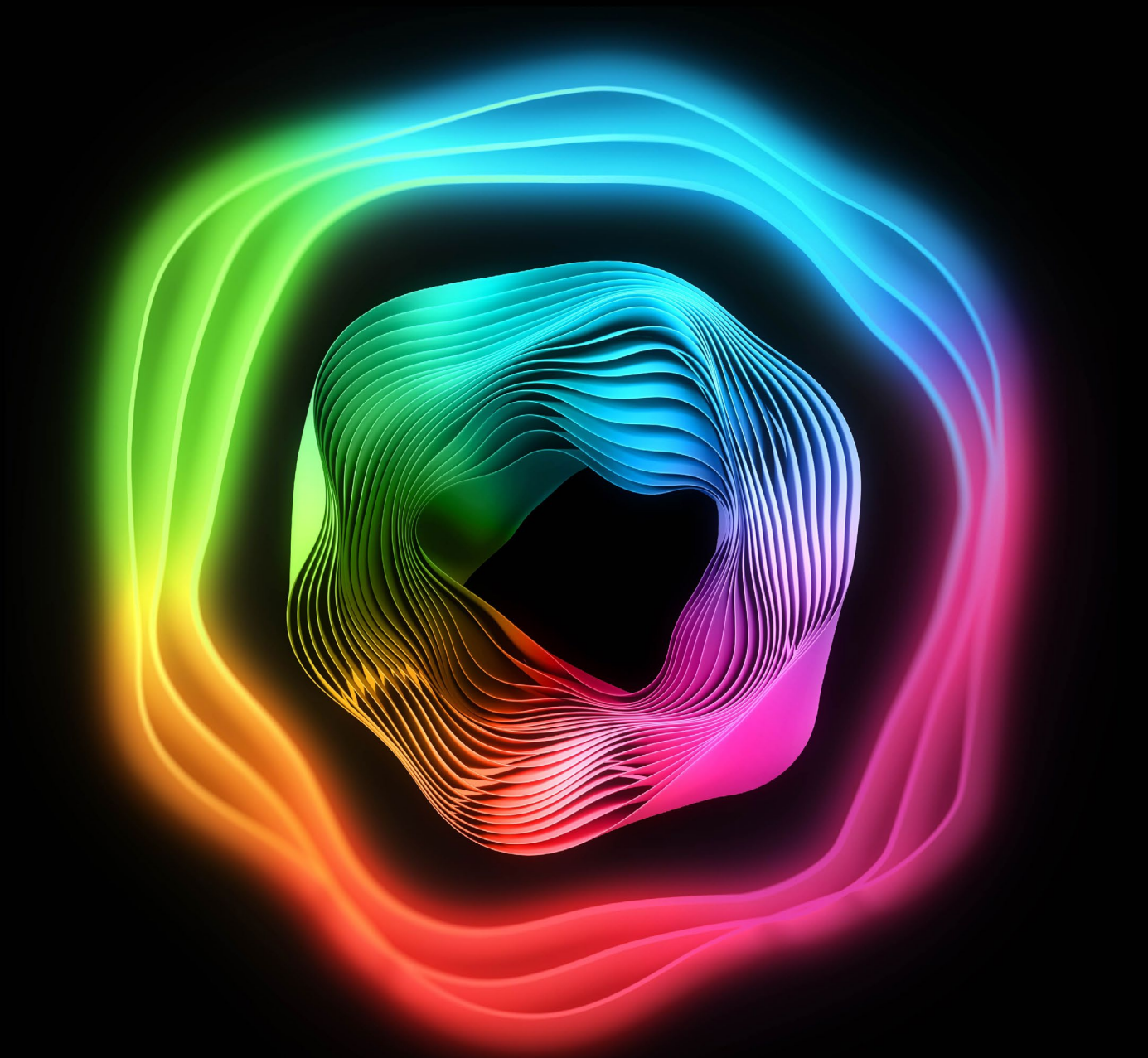
Deloitte.

The Call for AI Governance



MAKING AN IMPACT THAT MATTERS since 1845

The race is on! Companies increasingly maintain their competitive edge with the help of artificial intelligence. The proliferation of AI solutions demands professional management of quality, risks, and compliance. Good governance should enable as much as it controls – focused, lean, efficient and effective.

Introduction

Artificial Intelligence has transformed business operations around the globe, yielding a distinct productivity dividend, particularly in countries quick to embrace the technology, such as in the US and in China. Generative AI and Agentic AI are set to take this transformational effect even further, enhancing person-to-machine interaction, streamlining processes, and delivering fully new capabilities to products and processes across various sectors. The discussion is no longer whether AI delivers business value, rather how much, by when and for what cost. The cost equation is nuanced, encompassing implementation & maintenance effort, infrastructure, cost of compliance, and of non-compliance (penalties). Perhaps the greatest costs associated with AI are the costs of poor quality (development) or operation (deployment). They can result in substantial rework, recalls, business continuity risk, and reputational damage. The best defense against these come in the form of clear AI governance structures, robust AI quality management systems (QMS), and comprehensive AI risk management systems (RMS), efficiently implemented via digital platforms.

AI has become mainstream. It has long since progressed from data scientists, developers and enthusiasts to boardrooms, taking center stage in company strategies around the world. It is as flexible as it is potent. Yet it is not immune to error or naïve application. To capture the most value from AI, organizations must swiftly evaluate business cases, ensure quality implementation and master its risks – all features of sound governance. Part of a highly competitive market, made only more so with AI, they need AI Governance not only to be effective, but efficient – operationalized, systematized, collaborative, and immediate.

The Promise of AI, GenAI and Agentic AI

Artificial Intelligence, in its myriad forms — from Machine Learning to Generative AI and "agentic" AI— is a cornerstone of modern technological advancement. These systems can analyze vast datasets, identify patterns, and make predictions with remarkable accuracy. This capability positions AI as an invaluable asset in sectors ranging from finance to healthcare, logistics to telecommunications, retail to entertainment.

Generative AI, exemplified by models such as ChatGPT, has revolutionized human-computer interaction. These models process natural language and generate human-like responses, making advanced AI capabilities more accessible. This enables a wide range of applications, from content creation to customer service enhancement.

AI agents further extend these capabilities by performing tasks autonomously, learning from interactions, and adapting to new situations. This enables businesses to automate routine tasks, enhance decision-making processes, and foster innovation. Precisely its ability to execute and to use tools raises the stakes on both sides of the equation, in terms of both efficacy and risk.

Any technology is accompanied by risk. AI is no exception to this rule. Common IT risks such as data breaches, system failures, and cybersecurity threats are compounded by AI-specific challenges. Machine Learning models can scale and perpetuate biases present in training data. Generative AI systems might mislead with unreliable – however convincing – outputs. The autonomous decision-making capabilities of AI agents could introduce security risks, violate organizational or regulatory requirements without their human managers knowing. Subsequently existing and new regulatory requirements, such as the EU AI Act need to be operationalized and adhered to, without beleaguering organizations with burdensome red-tape that would strip away many of the benefits of utilizing AI.

New Opportunities and Challenges from "Big AI"

The contribution of foundation models is only the latest in a long trend of "democratizing AI" – starting with data made publicly available, then open-source algorithms, then code-collaboration tools such as GitHub. Yet foundation models may have had an even more profound impact, escaping the confines of the data science and programmer communities, putting the power of AI into the hands of less-advanced developers or even those with no coding expertise. The result is a "Cambrian explosion" of applications built on the core GenAI capabilities… and a dilemma to leaders. On the one hand, ease of access and application promise to usher in needed modernization to an AI-fueled business. On the other hand, it represents controllership challenges, where leaders may find themselves responsible for potentially harmful outcomes of AI-enabled systems, processes, and tools they may not even know exist.

This is not an entirely new challenge. The discipline of "Model Risk Management" (MRM) has existed for decades among top-tier banks, insurers, payment institutes, and others that rely heavily on models to power their business models, particularly in regulated sectors. However, in the past, most models were commissioned, developed, delivered and maintained through centralized functions of the organization. Models were often large, cumbersome models requiring a wealth of domain expertise and IT savvy to build; it would be unimaginable to attempt such projects "under the radar" without the formal support of expert teams throughout the organization. Constructing such models was labor intensive, requiring significant budgets. Even updating them was too costly to perform more than once yearly.

That changed already with AI. High-performing classifier models could be developed by a single data scientist – armed with enough reliable data – within a matter of weeks rather than months… and updated in days. The foundation models of GenAI have made this creative process even easier. Whether using LLMs to extract targeted information from documents or coding assistants to speed the development and documentation of the next generation of model, generative AI introduced new modeling possibilities and efficiencies, opening them up to a broader array of use cases than "credit risk" or "pricing" and activating a wider community to develop models or AI systems on their own. The new, decentralized dynamic has led to an abundance of creativity and an adoption of the technology. It has also led to proliferation of models and systems distributed throughout the organization, losing oversight, let alone control.

Regulatory Requirements: The AI Act

The regulatory landscape is evolving, particularly in Europe with the introduction of the AI Act, the world's first wide-reaching regulation of AI. The AI Act is a product safety regulation aimed at protecting EU citizens from potential harm of ill-conceived of poorly implemented AI. It centers around ethical and quality principles – responsibility, security, transparency, resilience, fairness, ecology, human oversight. Further, it mandates sound governance practices, especially for systems classified as High-Risk AI.

Risk Categorization

The AI Act does not hold all AI systems to the same standard. The absolute first step in the regulatory process is to categorize AI models and systems, performed differently depending on the nature of the AI:

- Single Purpose AI systems (SPAI) – categorization along the ethical considerations of the use case, its impact on EU citizens and their fundamental rights.
- General Purpose AI models (GPAI) – because they are multi-purpose, GPAI-models (essentially the foundation models behind Generative AI) cannot be categorized by use case, but rather by the aggregate capabilities, measured by the training effort invested into creating the model

SPAI systems may be derived from machine learning on a limited dataset for an optimized case, or may make use of a GPAI model, tuning it for a particular use case.

The SPAI can fall into any of three categories … and one "transversal" sub-category:

1. Unacceptable Risk – Forbidden

2. High Risk (HRAI) – Regulated

3. Non-High Risk (NHRAI) – not Regulated

The details of which use cases are associated with which category are explained on the Deloitte EU AI Act site, in the AI Act Summary, and in detail in the Deep Dive to the AI Act. The "transversal" sub-category refers to "transparency obligations" by which the deployer of an AI system (any SPAI system or GPAI model) must duly inform end-users that they are indeed interacting with an AI, not with a human being. These include, but are not limited to, use cases such as chatbots/voicebots or deep fake images/audio/video.

The focus of the AI act is clearly set on high-risk systems – HRAI. For these, the regulation requires two critical components of this governance framework:

1. the Quality Management System (QMS) and

2. the Risk Management System (RMS)

The Quality Management System (QMS)
Building to quality standards from the start is also the smartest strategy to avoid costs of poor quality, growing exponentially the later they are found in the value chain. Beyond rework, quality defects may inflict damage to the perception of the brand and loyalty of customers. The risk is not confined to startups and smaller business, nor is it unique to AI. Some of the strongest brands in the market can be taken by surprise – with engine defects (automotive), overheating batteries (electronics), chemical pollutants (food).

The QMS establishes procedures by which products are built and tested to a given standard. It starts by clearly defining critical-to-quality requirements, translating these into a technical specification, identifying necessary skills for developers and testers, then staffing appropriately, and delineating clear areas of responsibility (segregation of duties). It puts mechanisms in place to prevent potentially costly shortcuts, involving multiple stakeholders and providing a safe environment to raise issues. The primary focus of the QMS is to avoid defects by building quality directly in from the start. However, there is likely not a software in existence whose defects were all known and resolved prior to launch – especially not in the age of hackers and cyber-crime. To that end, the secondary focus of the QMS is to remain vigilant throughout the product lifecycle, acting quickly to patch defects found only post-deployment. Acting quickly means having anticipated to some degree that issues can arise and how best to organize in response to them. The failure mode & effects analysis (FMEA) is a useful exercise to include within the QMS to enable the requisite agility when issues do arise. Some of the contingency plans arising from the FMEA may be actions to be executed by the provider, others by the deployer. The latter are part of the QMS requirement that provider deliver "instructions for use" of the AI to the deployer.

The AI Act requires a risk categorization of all AI systems, followed by specific requirements for high-risk AI systems: technical documentation, the Conformity Assessment, the Declaration of Conformity, and – if required – the Fundamental Rights Assessment, to formalize the QMS process. It does not specify what is meant by technical documentation, however most practitioners will understand this as:

- business requirements
- technical specifications
- data lineage
- model cards/factsheets
- test cases & testing results (unit tests, system tests, system integration tests, user acceptance tests)
- failure mode & effects analysis (FMEA)
- risk classification
- developer documentation
- user documentation

The QMS is particularly relevant to AI providers—those developing AI systems. It emphasizes the need for comprehensive technical documentation, stringent approval and testing processes, conformity assessment, and the Declaration of Conformity. The importance of AI quality cannot be overstated. High-quality AI systems ensure operational efficiency, minimize costly rework, and safeguard against business continuity risks. They prevent the leakage of confidential information, protect the organization's reputation, and maintain customer trust. Moreover, they help avoid potential compliance breaches that could result in legal repercussions and financial penalties.

As providers (developers) of AI have the most influence over the quality of AI systems, they play the central role in the QMS. Providers must ensure that AI systems are built to high standards and can perform as intended across various environments. This involves maintaining and updating the systems as necessary and addressing any issues that arise post-deployment with agility and precision. For instance, a financial institution using a Machine Learning model for credit scoring must rigorously test the model to ensure it accurately assesses risk without bias, thereby upholding regulatory compliance and customer trust.

The Risk Management System (RMS)
Monitoring, issue logging and resolution are the centerpiece to the RMS. Keeping an eye on early warning indicators enables providers to identify and assess potential issues before they become costly problems. Issue logging provides the backbone for organized feedback to providers; issue resolution the rigor to enforce timely response. They are essential ingredients to any business continuity policy, for instance dictating whether a system may be allowed to continue in operation or should be temporarily taken offline until a patch is available. They are also regulatory requirements of the AI Act for high-risk AI systems – although they are equally relevant to all systems in operation. Closely related is cyber threat monitoring. While it can and should be integrated into AI Governance, it is also part of a wider cyber security framework – encompassing access privileges, threat surface analysis, identification and selection of suitable cyber defense tools, rigorous penetration testing/red-teaming. Cyber security considerations address both vulnerabilities to systems in general as well as those particular to AI, such as adversarial attacks or prompt injections.

The RMS is crucial for AI deployers—those who implement and use AI systems. It focuses on monitoring the AI system's performance, logging issues, and reporting any malfunctions to the EU public database. Effective AI risk management ensures that AI systems continue to operate as intended over time. It enables organizations to identify potential issues before they materialize in the market, thereby preempting any significant disruptions or adverse impacts. By pinpointing the root causes of issues, organizations can swiftly resolve them, maintaining operational integrity and user trust.

Where **providers** determine how an AI system is constructed, they have little direct influence on how, where and for whom it is used. Only **deployers** of AI are in a position to monitor the AI system that they operate, the key function of the RMS. They must establish robust monitoring frameworks, capable of detecting anomalies, performance drifts, or any unintended outcomes. For example, an e-commerce company using AI for personalized recommendations must continuously monitor the system to ensure it adapts to changing customer preferences and market trends, preventing any decline in user engagement or satisfaction.

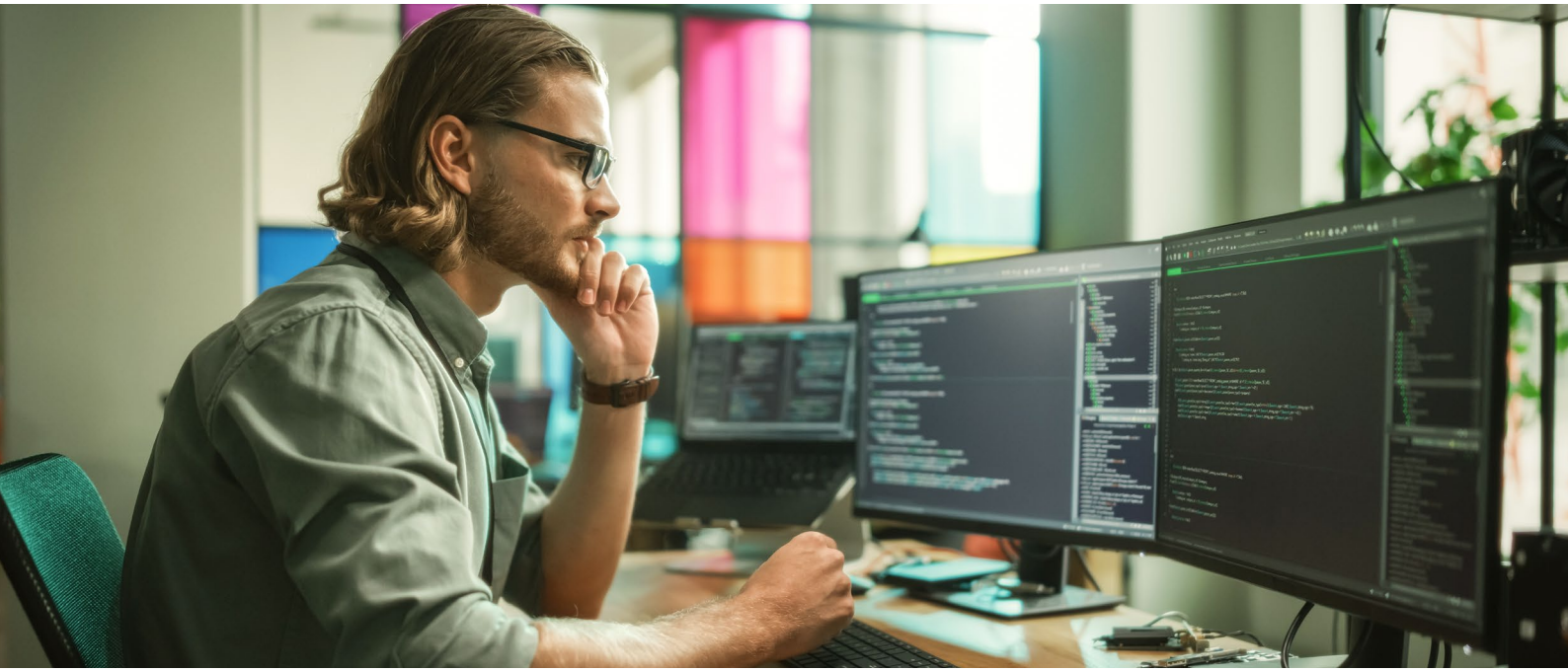
Monitoring at scale is unrealistic without a digital platform to automate the data collection, evaluation, drill-down and visualization. Such a platform enables organizations to automate governance workflows, track compliance metrics, and maintain comprehensive records of all AI-related activities. This reduces the administrative burden and enhances the accuracy and consistency of governance practices, providing a higher level of transparency and accountability in AI operations.

The Notion of AI Governance

The QMS and RMS form parts of a continuum, complementing each other and defining the respective roles and responsibilities of the provider vs the deployer. Simplifying the picture by omitting the facilitating roles of importers and distributors the interplay of the QMS and the RMS form a logical process with distinct ownership and handoffs:

	Concept	Build	Deployment	Operation	Assurance
Provider (QMS)	x	x			x
Deployer (RMS)			x	x	

Tabel 1 *The QMS and RMS form parts of an governance value chain along the lifecycle of an AI system.*



Together, the QMS and RMS provide an effective cross-organizational governance structure to manage the AI lifecycle, ensuring the AI continues to function as intended until decommissioning. Quality and risk management are important to any product – and they are even important for AI-systems due to their probabilistic nature. Unlike the "syntax errors" of rules-based systems, an AI system will always output an answer, at varying degrees of statistical confidence. The fact that they are trained, not programmed with rules, means that they can fall victim to "data drift" and "concept shift" if not designed to be sufficiently robust. Put simply: the world may have moved on since the time when the model was trained, meaning its training data, and thus its model weights, are no longer reflective of the data observed in the operational environment. It is a perfect example of how RMS and QMS work together: the RMS detects the drift, the QMS resolves it with subsequent retraining.

The QMS and the RMS are essential components of sound governance. Yet governance is a wider topic than quality and risk management. Governance encompasses organizational structures, systems, practices and processes to enable management to fulfill its duties to ensure quality, manage risk, enforce accountability and fulfill compliance obligations. It begins with policies & procedures, extends to decision-making bodies & approval processes, through to review boards and controls.

For AI governance to be effective, it must cover a wide spectrum of processes that fit the particular organization, its products & services, the demands of its industry. Its components can be viewed across four pillars:

Structures	Practices	Processes	Systems
• Steering Committee (Ethics, Investment, Quality, Risk)	• Oversight & accountability (QMS, RMS)	• Overview/inventory	• Operationalizing the practices, automating the processes
• Advising AI CoE	• Qualifications & skills	• Approval process & quality gates (QMS)	• Workflows (QMS)
• Internal Audit	• Empowerment & responsibility	• Escalation & issue review	• Technical dcumentation (QMS)
• Training/Upskilling Facility	• Developer guardrails (QMS)	• Risk assessment/classification	• Model monitoring over time, triggers & alerts (RMS)
• Distinct roles & responsibilities	• Business continuity & risk mitigation strategies (FMEA)	• Testing/Model Validation	• Issue logging and resolution (RMS)
• Relevant stakeholders	• Communication (Awareness)	• Compliance	• Single source of truth, serving information needs at various levels
		• Change requests/issue resolution	

Tabel 2 *Governance is multi-dimensional and multi-disciplinary, spanning people, process and technology.*

For AI governance to be effective, it must also be efficient. Anything less will negate the positive productivity impact brought about by AI, and likely not survive in a cost-conscious business environment. It must be highly automated in order to be nimble and responsive in the fast-paced AI technology scene. This is the role of the "Systems" pillar – to operationalize governance: automating controls, accelerating decision-making, easing collaboration and facilitating management oversight duties. Fundamental to a governance platform is that it integrate seamlessly with other systems. The RMS presents a perfect example. Risk monitoring can only function efficiently if directly linked (via API, application programming interface) to the AI models it is tasked to monitor. The governance platform must be flexible to interface with those AI systems on any cloud or on-premise configuration, as most organizations adopt a nuanced, hybrid, multi-cloud strategy for their AI systems. They seek to avoid vendor lock-in, to profit from particular strengths of any cloud and GenAI provider for the use case in question, and opt to keep some, particularly sensitive AI systems (machine learning models trained and applied on sensitive data) entirely within their own four walls.

The Road to Good Governance

The transition of AI tools and systems from Proof of Concept (PoC) to full-scale production reveals the complexities in managing AI quality, AI risk and AI compliance. As AI-powered systems become integral to business operations, both practitioners and senior management recognize the necessity of managing the AI lifecycle and its associated risks comprehensively.

Implementing robust, efficient governance structures is crucial for overseeing the development, deployment, and operation of AI systems. Effective governance ensures that AI systems perform reliably and ethically, aligning with organizational goals throughout their lifecycle. This requires establishing and maintaining clear frameworks and control management practices. Identifying and mastering AI risks is a central obligations of management. It ensures that AI systems can be implemented with confidence, knowing that there are structures in place to manage any potential risks effectively.

The Committee of Sponsoring Organizations of the Treadway Commission's Internal Control-Integrated Framework (COSO-ICIF) has long been the golden standard for designing and implementing effective systems of internal control. As organizations increasingly adopt AI, it seems appropriate to align AI governance principles with the COSO-ICIF. The COSO framework addresses designing and implementing effective governance systems, in five components, i.e. Control Environment, Risk Assessment, Control Activities, Information and Communication, and Monitoring Activities. AI Governance should embrace COSO-ICIF's components and also reflect the organizational and technological requirements associated with adopting AI in modern organizations.

Enabling AI Governance – how to get there is an interplay of five key contributors:

1. Vision and Strategy,

2. People and Culture,

3. Organizational Structure,

4. Process and Controls, and

5. Technology.

Vision and Strategy
The overarching goals behind the implementation of artificial intelligence within an organization are anchored in its strategy. Its careful construction and clear articulation provide organizations with the orientation they need to align efforts in leveraging AI technology to enable growth, strengthen the competitive edge and thereby the future prospects of the organization – without exposing themselves to undue risks.

Organizational Structure
Mobilizing an organization to execute on its AI vision and strategy in a coordinated manner starts with the right structure – fit for purpose in consideration of both the objectives and the environment in which the organization acts. The design and distribution of roles, responsibilities, and reporting lines lay the foundation for effective implementation, quality control and risk management. Organizational structures can produce vastly different dynamics – from regimental hierarchies focused on execution … to fluid, cross-functional "tribes" ("cells" or "CoEs") focused on innovation, a construct arguably more suitable to efficient, safe implementation and adoption of AI. Steering committees of senior executives and subject matter experts combined, each with individual roles and responsibilities, collectively provide oversight and strategic direction to these teams.

People and Culture
Governance is only as effective as the individuals who implement it. Ensuring that those responsible for governance have the necessary qualifications, education, awareness, and sensitivity to the implications of AI risks is paramount. Clear roles, communication channels, and metrics are essential for a coordinated effort.

Training programs and continuous learning opportunities can enhance the competency of governance teams, ensuring they are well-equipped to handle the complexities of AI risk management. By fostering a culture of accountability and transparency, organizations can build a robust foundation for AI governance.

Processes & Controls
Processes provide the structure needed to bring order from chaos. They ensure that organizational structures are utilized effectively and guide the organization to fulfill its purpose. Without well-defined processes, even the most qualified individuals may struggle to achieve consistent results.

Standardized procedures for AI development, deployment, and monitoring can streamline governance activities, reduce variability, and enhance the predictability of AI systems. By establishing straightforward guidelines and protocols, organizations can ensure that all AI-related tasks are performed consistently and in accordance with best practices.

Exemplary AI Governance-Risks and related controls that companies need to consider are:

Risk	Control
Questionable use cases out-of-sync with company values or stated purpose	Ethics incorporated into the AI governance framework, on the agenda of steering committees/decision-makers
Poor quality implementation (accuracy, bias, privacy, security, instability)	Guidelines and developer guardrails built on Trustworthy AI principles, best associated with quantitative standards
Loss of quality or viability between prototype and production	Rigorous testing procedures: (automated) unit tests, integration tests, system tests, user acceptance tests
Performance degradation, drift or instability over time after being deployed into production	Active monitoring of AI systems in production aligned to highly relevant, case-specific metrics and control limits (thresholds, corridors)
Business interruption or similar risk, lacking contingency plans in case of malfunction	Anticipatory analysis of potential failure modes and limited/directional planning of emergency measures (incl. “kill-switches”) to be taken in event of failure

Tabel 3 *An overview of general operational risks and associated control strategies.*

Technology
Former US Deputy Attorney General Paul McNulty once famously said, "If you think compliance is expensive, try non-compliance!" His message was abundantly clear. Nevertheless, maintaining good governance and upholding compliance does indeed weigh on the ongoing profitability of an organization. The burden can be significant, especially for small or midsize companies in today's highly competitive market, made only more competitive by AI. Compliance is important, but it must be made efficient. An essential feature of sustainable, good governance is that it automates where possible, supported by technologies, to achieve efficiency – whether for purposes of enterprise risk management or regulatory compliance.

For organizations managing multiple AI tools and systems, the tasks associated with QMS and RMS can become overwhelming if handled manually. Therefore, digitized processes are essential. An efficient governance platform operationalizes the organization’s governance policies and practices, streamlining these processes and making them more manageable and effective.

Operationalizing AI Governance

Deloitte's AI Quality & Risk Management solution addresses the need for robust governance structures. This solution enables sound Enterprise Risk Management and fulfills both the QMS and RMS requirements of the AI Act in an efficient and scalable manner.

By leveraging Deloitte's extensive expertise and IBM's WatsonX.Governance platform, this solution offers a comprehensive framework for managing AI risks. It provides a holistic view of AI governance, integrating various tools and processes into a single, intuitive platform. This enables organizations to monitor AI systems in real-time, ensure compliance with regulatory requirements, and respond promptly to any issues that arise.

One of the distinguishing features of Deloitte's AI Quality & Risk Management is its ability to provide detailed analytics and insights into AI system performance. By utilizing advanced monitoring tools, organizations can gain a deeper understanding of how their AI systems operate, identify potential risks early, and take proactive measures to mitigate these risks. Additionally, the solution offers automated compliance tracking, ensuring that all regulatory requirements are met without imposing significant administrative burdens on the organization.

Furthermore, Deloitte's and IBM's solution is designed to be highly scalable and adaptable, making it suitable for organizations of all sizes and across various industries. Whether an organization is just beginning its AI journey or has already implemented several AI systems, the solution can be tailored to meet its specific needs and objectives.

AI Ethics: Doing the Right Thing

While AI is a technical topic, it is one with widespread ramifications. Yet it is seldom the technology itself which is dangerous, rather the intent of the people who put it to use. This has been true since the dawn of history; while fire can be used productively to refine ore, it can also cause harm – from recklessness or bad intent. Automobiles and airplanes connect people and places, yet their misuse can also end in tragedy.

This begs the question of how human intent and requisite care – or lack thereof – may impact a revolutionary and multi-use technology such as AI in its various forms. Leaders ultimately accountable for the application of AI must ask questions such as:

- Who stands to gain and who may lose from the application?
- Does it fit to the values of our organization, to our collective purpose?
- What approval and oversight mechanisms should we put in place?
- How would the application affect our business model, our employees?
- How would not implementing it affect our business?

These and other questions – and the answers to them – will distinguish the leaders from the rest. Again, ethics are not only the compass for AI Governance – or any product management – they can also be found at the heart of regulation such as the AI Act.

AI Quality: Holistic Performance Evaluation

Transparency, robustness, impartiality, representativeness, confidentiality, and security are vital dimensions in evaluating AI quality. Each of these dimensions requires meticulous measurement to ensure AI systems operate reliably and ethically.

- Transparency involves making the AI's decision-making process understandable to stakeholders. Metrics such as explainability scores assess how easily humans can interpret AI decisions. For example, in healthcare, an explainable AI model can reveal why it recommends a particular treatment, thus building trust among doctors and patients.
- Robustness refers to an AI system's ability to perform reliably under various conditions. Stress testing metrics, which involve subjecting AI to unexpected inputs, can evaluate robustness. In autonomous vehicles, robustness ensures the AI navigates safely in diverse weather conditions.
- Impartiality and representativeness are crucial to avoiding biases in AI decisions. Fairness metrics, such as disparate impact analysis, measure whether AI outcomes are equitable across different demographic groups. For instance, in hiring processes, ensuring that an AI tool evaluates candidates without bias helps maintain diversity and inclusion.
- Confidentiality and security ensure that AI systems protect sensitive data against unauthorized access and breaches. Metrics like encryption strength and vulnerability assessment scores are essential. In financial services, robust security measures prevent fraud and protect customer information.

AI Risk Management: Everything under Control

Where AI Quality is essential to the development process, the focus shifts to AI Risk Management after deployment. AI Quality seeks to avoid quality defects – through a combination of standards, guidelines and guardrails. AI Risk Management seeks to monitor performance, identifying and remediating issues before they grow into business problems.

Sound AI Risk Management starts long before deployment, long before monitoring goes live. Where the governance contribution of the QMS may be generally applicable to any AI in development, it is a very different story for the governance role of the RMS for AI systems in production. Each AI system will have its own function and be evaluated on different criteria. Fairness and privacy may be critical for an algorithmic credit risk classifier, deciding to approve or reject loan applications. It will not be important for readings emitted by a sensor on an assembly line (although related concepts of confidentiality and representativeness may indeed be relevant). Other concepts may be of more universal interest, such as robustness – whether an AI system continues to operate as intended when exposed to imperfect data, noise, even resilient against adversarial attack. Choosing the right metrics forms the basis of AI risk management, followed by identifying a safe performance corridor and then by monitoring the AI systems KPI outputs over time to ensure continued operation within that corridor.

Another key component to risk management is identification, logging, and resolution of issues. Again, setting up logging starts long before go-live, starting with failure modes and effects analysis to anticipate potential breakdowns and have strategies in place for business continuity and swift issue resolution.

Once AI systems have passed rigorous testing and are deployed into production, the active monitoring and issue logging begins. For a handful of models, this can be achieved to some degree through ad-hoc solutions built in-house. Beyond that, it is strongly recommended to graduate to the level of sophistication and responsiveness provided by a digital governance platform: robust, scalable, responsive, intuitive.

Tangible Benefits

Well implemented, sound AI Governance structures balance strategic aims of harnessing the potential of AI technologies, fostering an environment that drives innovation without compromises. Sound AI Governance enforces responsibility and ethical business decisions. It integrates AI Quality principles of safety, reliability, transparency, fairness, and confidentiality. Sound AI Governance also focuses attention anticipating and managing risks, should they materialize, by tracking performance, leaving an audit trail of accountability and benchmark for continuous improvement.

Economically speaking, the investment in sound AI Governance pays dividends by avoiding the cost of quality defects, avoiding inaccurate decisions, business interruptions, recalls & product liabilities, compliance penalties, reputational damage or a deteriorating client base. Effective governance structures enhance organizational reputation, build stakeholder trust, and create a stable environment for further investment in AI technologies.

Moreover, effective AI governance fulfills obligations of management oversight towards all stakeholders – among them customers, employees, supervisors. By ensuring that issues are promptly identified and addressed, organizations can maintain high standards of service and employee satisfaction, a virtuous circle confronts issues head-on, learns from them, and integrates learnings into the next iteration of products and services.

The Bottom Line

The promise of AI, particularly Generative and Agentic AI, is immense. However, realizing these benefits requires a balanced approach that acknowledges and manages the associated risks. Robust governance structures, including a well-defined QMS and RMS, are essential for ensuring that AI systems perform reliably and ethically throughout their lifecycle.

Deloitte's AI Quality & Risk Management solution offers a comprehensive framework for AI Governance, addressing both regulatory requirements and practical challenges. By focusing on People, Process, and Technology, organizations can implement AI with confidence, harnessing its potential while mitigating its risks.

The future with AI is bright, but not without pitfalls. Success requires both pioneering spirit and careful stewardship. AI Governance, guided by ethics, with a mindset of unrelenting quality and efficient risk management practices, is a valuable tool helping leadership navigate the complexities of AI implementation, enjoying high-performance and building trust in their journeys to become truly AI-fueled organizations.

Your contacts

David Thogmartin
Partner
aiStudio
AI & Data Analytics | Risk Advisory
Tel: +49 211 8772 2336
dthogmartin@deloitte.de

Jan Grüne
Director
Digital Internal Audit & Data Intelligence
Tel: +49 69 75695 6222
jgruene@deloitte.de

Torsten Berge
Director
Business Assurance
Tel: +49 151 58072499
tberge@deloitte.de

Deloitte.

Issue 01/2025