Charter of Trust

AI Policy Paper

February 2026

Contributors:

- Christoph Peylo, Bosch - AI Working Group Co-Chair
- Sonja Zillner, Siemens - AI Working Group Co-Chair
- Ki Hyun Park, Mitsubishi Heavy Industries
- Linda Strick, Cloud Security Alliance
- Michael Moore, Atos
- Richard Skalt, TÜV SÜD
- Suzy Button, Elastic
- Tatiana Bellendir, Siemens
- Edoardo Mancini, Charter of Trust

Classification CoT Public

Executive Summary

Artificial intelligence has become a critical component of modern industrial processes, cybersecurity operations, and digital infrastructure.

As companies increasingly build and integrate their own AI capabilities, **the need for secure, trustworthy, and compliant digital environments has never been more pressing**. This paper provides a clear framework for organisations to navigate this landscape, marked by **concentrated provider ecosystems, fragmented global regulations, and geopolitical supply-chain risks**, alongside the internal requirements necessary to build AI responsibly.

This paper assumes that readers are already familiar with advanced AI system architectures, security governance, and regulatory frameworks. Rather than reiterating foundational concepts, it focuses on the structural, organisational, and geopolitical challenges that emerge once AI systems move into mission-critical and regulated environments.

A central focus of the paper is **helping organisations prepare for the EU AI Act**, based on the overarching principle that compliance cannot be treated as a simple checklist exercise, and should instead drive strategic transformation. Organisations are encouraged to ensure visibility over all AI systems in use, promoting alignment across technical, legal, and business functions. Strengthening governance is equally critical. Executive-level oversight, supported by operational teams, should lead to consistent, iterative risk assessment throughout the AI lifecycle, ensuring that performance, ethical, legal, and operational risks are identified and addressed early.

Quality management and rigorous documentation practices emerge as core enablers of compliance. By aligning with globally recognised AI-specific quality frameworks (e.g. ISO/IEC 42001 standards), organisations can build robust procedures for data governance, model validation, bias detection, monitoring, and corrective action. At the same time, audit-ready documentation systems can strengthen traceability and provide the evidence regulators expect.

Companies must also balance compliance investments against the financial and reputational risks of non-compliance. The penalties under the AI Act are substantial. However, proactive preparation not only reduces exposure, it can also create a competitive advantage by enabling faster innovation, strengthening customer relationships and regulator trust, and reducing uncertainty in product development.

Looking ahead, **organisations should treat AI governance as a long-term, adaptive discipline**. Regulation and technology will continue evolving, and resilience depends on flexible policies, modular system architectures, and scalable governance processes. Continuous monitoring of regulatory developments, active participation in standards-setting activities, and sustained investment in skills are essential to fostering a responsible AI culture centred around a holistic understanding of compliance.

Content

The Charter of Trust: Our Mission

Countries of establishment of CoT Partners/Associated Partners

On top of their countries of establishment, CoT Partners are also active in close to 170 countries worldwide.

Amidst an increasingly severe and complex threat landscape, the Charter of Trust (CoT) was established at the Munich Security Conference on 16 February 2018 as a non-profit alliance of leading global companies and organizations. Since then, a continuously evolving group of members and partners works together across sectors to strengthen cybersecurity, cultivate digital trust and make the digital world of tomorrow a safer place. Today, our initiative consists of 13 Partners and 17 Associated Partners operating in nearly 170 countries across five continents and representing more than 1.8 million employees.

All members endorse the **ten fundamental principles of CoT designed** to achieve **three overarching objectives:**

- To **protect** the data of individuals and companies;
- To **prevent** damage to people, companies, and infrastructure;
- To **create** a reliable foundation on which confidence in a networked, digital world can take root and grow.

Guided by these principles, the Charter of Trust is working to protect our increasingly digitized world and build a reliable foundation on which trust and digital innovation can flourish. It advances effective cybersecurity policies worldwide and offers expertise in areas such as artificial intelligence, post quantum cryptography, security by default, supply chain protection and education.

This publication is issued by the Charter of Trust's AI Working Group.

The AI Working Group

Artificial intelligence is transforming the cybersecurity landscape for businesses, offering both unprecedented opportunities and complex challenges. On the positive side, AI enhances cybersecurity by automating threat detection, improving response times, and predicting potential security breaches through advanced analytics and machine learning algorithms. These capabilities enable businesses to proactively defend against cyber threats, minimize vulnerabilities, and enhance overall security posture.

However, **the integration of AI also introduces new cybersecurity risks**. AI systems can themselves become targets for cyberattacks, potentially being manipulated or exploited by malicious actors. Additionally, the complexity and opacity of some AI algorithms can make it difficult to identify and mitigate biases and vulnerabilities, leading to security gaps.

In light of the regulatory advancements on AI, **the mission of Charter of Trust's AI Working Group is to provide clear guidelines for ensuring innovative but also secure and compliant with regulatory requirements.**

Disclaimer

The following document serves as an overview and general information resource only. It is not intended to provide legal advice or guidance of any kind. While efforts have been made to ensure the accuracy and completeness of the information presented herein, it may not encompass all legal nuances or variations applicable to specific circumstances.

Readers are encouraged to consult with qualified legal professionals or advisors regarding their particular situations or concerns. Reliance solely on the information contained in this document is done at the reader's own risk. The author and publisher disclaim any liability for any loss or damage arising directly or indirectly from the use of or reliance on this document.

1. Introduction

Artificial intelligence has moved from experimental technology to an essential component of modern industrial processes, public services, and digital infrastructures. As companies and organisations increasingly build their own AI systems, the need for secure, trustworthy, and compliant development practices becomes critical. AI now influences decisions in manufacturing, mobility, healthcare, finance, energy, and public administration—domains where errors, unintended behaviour, or malicious manipulation can have significant safety, economic, and societal consequences. Ensuring the security and compliance of AI systems is therefore not optional; it is a prerequisite for their responsible use in mission-critical environments.

Building AI systems introduces unique risks that differ fundamentally from those associated with traditional software. While these risks are widely discussed, they are still frequently underestimated in operational environments. Many organisations discover only after deployment that data lineage is unclear, model behaviour is insufficiently documented, or accountability for AI-driven decisions has not been formally defined. At that point, technical excellence alone no longer compensates for governance debt.

For the Charter of Trust and the Munich Security Conference community, secure and trustworthy AI is a central policy concern. Both forums bring together leaders from industry, government, and civil society who shape global security norms and technological governance. As AI systems increasingly influence geopolitical stability, economic resilience, and critical infrastructure protection, security policy must reflect the challenges and opportunities introduced by advanced AI technologies. This paper therefore provides guidance not only for technical practitioners, but also for decision-makers responsible for regulatory alignment, security strategy, and international cooperation.

The work presented here builds on several years of collaboration within the Charter of Trust AI Working Group. Earlier publications examined foundational principles for trustworthy AI, emerging regulatory frameworks, and implications for cybersecurity. This report expands those efforts by offering practical guidance for organisations that develop their own AI systems, complementing prior analyses on governance and risk. It continues a timeline of work that includes the group's contributions to responsible AI standards, alignment with EU regulatory developments, and cross-industry dialogue facilitated through the Munich Security Conference.

2. External Situation for Companies building AI

Artificial intelligence has moved from experimental technology to core business infrastructure, but the external landscape for AI development and deployment has grown increasingly complex. Organizations building or integrating AI systems now operate within a multi-dimensional environment shaped by concentrated provider ecosystems, diverging regulatory frameworks, and geopolitical supply chain constraints.

This chapter maps the critical external factors that enterprises must navigate when implementing AI solutions.

2.1 Global Situation of AI providers

AI capability is concentrated in a few technology ecosystems with distinct benefits — and dependencies:

- **United States** leads at the model and platform level through OpenAI, Microsoft (Azure OpenAI), Google/DeepMind, Anthropic, Meta, and Amazon (Bedrock). Azure OpenAI Service and AWS Bedrock dominate enterprise distribution, with Azure emphasizing data residency and AWS expanding into agent orchestration. NVIDIA controls training and inference silicon, creating critical concentration risk across the stack.
- **Europe's** primary independent developers are Mistral (France), Aleph Alpha (Germany), and Black Forest Labs (Germany). Mistral combines open-weight and proprietary models with strategic backing, while Aleph Alpha targets sovereign, auditable AI for regulated sectors. Meta's Llama and Hugging Face provide additional options to avoid vendor lock-in.
- **China's** platforms—Baidu (ERNIE), Alibaba (Qwen), Tencent, ByteDance—iterate rapidly under content-control regimes, creating parallel supply chains for Western multinationals operating there.
- **India** is emerging through Sarvam AI and Krutrim, offering cost-efficient engineering and evolving data protection frameworks.
- **Japan** balances domestic champions (NTT's tsuzumi, NEC's cotomi, Fujitsu's Takane) with global platforms, providing sovereign options alongside frontier model access.

The global AI provider landscape is characterized by U.S. dominance in foundation models and compute infrastructure, Europe's push for sovereign alternatives, China's content-controlled platforms, and emerging hubs in India and Japan—each offering distinct capabilities, dependencies, and compliance obligations.

As a result, Vendor selection is inseparable from geopolitical and compliance strategy.

2.2 Fragmentation of International AI Legislation

The global regulatory landscape for artificial intelligence is highly fragmented. While the EU is moving forward with the AI Act and a risk-based, binding regulatory approach, other major jurisdictions such as the United States, China, India, Japan, and Canada are pursuing divergent strategies. This lack of harmonization creates significant compliance challenges for multinational organizations, which must navigate overlapping, and sometimes conflicting, legal requirements:

Region	Core Regulatory Drivers	Status / Key Obligations
EU	EU AI Act, GDPR, NIS2	Phased binding obligations through 2027; high-risk system controls; documentation and monitoring
U.S.	State laws + NIST AI RMF	Bias audits (NYC), impact assessments (CO 2026), automated decision transparency (CA)
China	Provider-centric generative AI rules	Mandatory security assessments, content controls, provenance verification
India	DPDP Act	Data protection baseline; advisory-driven responsibilities
Japan	AI Promotion Act, APPI + 2024 AI Guidelines	Lifecycle governance, transparency expectations
Canada	Privacy frameworks after AIDA pause	Automated decision-making guidance; evolving enforcement

In addition, **global AI related Standards**, such as ISO/IEC 42001 (AI management systems), ISO/IEC 23894 (risk management) or automotive safety norms (ISO 21448 SOTIF, ISO 26262) are emerging.

As regulatory frameworks are fragmented and not harmonized, multinational organizations must simultaneously satisfy multiple, sometimes conflicting, requirements.

As a consequence, organizations must design their AI governance and compliance strategies to address multiple, often inconsistent, regulatory regimes. This fragmentation increases the complexity and cost of compliance, and raises the risk of regulatory gaps or overlaps.

2.3 Supply Chain Risks and Sourcing Restriction

Beyond regulation, physical supply chains face concentration and dependency risks. AI supply chains are increasingly exposed to geopolitical tensions, export controls, and sourcing restrictions. Companies face risks related to the availability and reliability of critical components, such as advanced semiconductors, specialized hardware, and proprietary software. Export bans and trade restrictions—especially between major economies—can disrupt access to essential technologies and services.

Vendor concentration further amplifies these risks. Reliance on a small number of providers for cloud infrastructure, foundational models, or data processing can create single points of failure. Organizations must assess the resilience of their supply chains, diversify sourcing strategies, and monitor regulatory developments that may impact procurement or operational continuity.

Export controls on advanced semiconductors, concentration in model providers, and opaque training data provenance create vulnerabilities that governance frameworks must address. Documentation, once an afterthought, has become both a legal necessity and a competitive differentiator in demonstrating trustworthiness.

For enterprises, understanding this external terrain is no longer optional. Strategic decisions about which AI systems to build, which vendors to trust, where to source compute capacity, and how to structure governance depend on clear-eyed assessment of the provider landscape, regulatory obligations, supply chain exposure, and documentation expectations outlined in this chapter.

2.4 Conclusions

The external AI landscape demands that organizations move beyond tactical compliance toward strategic resilience. Success requires diversifying provider dependencies, building governance that spans fragmented regulatory regimes, treating documentation as core infrastructure, and integrating supply chain due diligence into procurement and risk management.

Enterprises that defer these decisions face escalating consequences: regulatory penalties under the EU AI Act and state laws, exposure to export control disruptions, and supply chain vulnerabilities. Conversely, organizations that establish transparent documentation, diversified suppliers, and adaptive governance frameworks position themselves not merely to comply, but to compete—building stakeholder trust and capturing opportunities that responsible AI creates.

The path forward requires treating AI governance as a continuous, cross-functional discipline. Organizations that invest now in mapping the provider landscape, understanding regulatory obligations, securing supply chains, and institutionalizing documentation will navigate this complexity with confidence.

3. Internal Situation for Companies that are building AI

Companies that choose to build their own AI systems face a complex set of internal challenges that go far beyond technical development. Successful AI development requires not only data, models, and infrastructure but also governance structures, organizational capabilities, and processes that ensure reliability, safety, and compliance. This chapter outlines the internal conditions that influence whether companies can build AI responsibly and effectively.

3.1 Motivation for Build Versus Buy

Organisations must decide whether AI components should be developed internally, procured from vendors, or implemented through hybrid approaches. Off-the-shelf solutions offer rapid deployment, lower initial cost, and predictable service levels. However, they often provide limited transparency, restricted customization, and challenges in meeting specific regulatory obligations or integration needs.

In contrast, building AI internally allows deeper control over data, model behaviour, deployment environments, and compliance documentation. It also enables organisations to embed proprietary knowledge into models and create solutions tailored to internal workflows. This approach, however, requires substantial investment in data quality, engineering capability, documentation processes, and long-term maintenance. Many companies therefore adopt hybrid models, combining commercial or open-source components with in-house fine-tuning and domain-specific logic. The key determinant is whether the organisation possesses the competencies and resources to manage the full lifecycle of an AI system.

Business Area	Typical Use Cases	Suitable AI Technology Families
Operations & Manufacturing	Predictive maintenance, process optimization, quality inspection	Classical machine learning, deep learning
Customer Support & Interaction	Virtual assistants, ticket triage, conversational services	Generative AI, agentic AI systems
Marketing & Sales	Personalization, lead scoring, content generation	Classical machine learning, generative AI
Product Design & R&D	Concept ideation, code generation, simulation support	Classical machine learning, reinforcement learning
Supply Chain & Management	Demand forecasting, routing, inventory optimization	Classical machine learning, reinforcement learning
HR & Talent Management	Planning, assessments, scenario modelling	Generative AI, symbolic AI
Compliance Support & Knowledge Management	Screening support, skill matching, training content	Agentic AI, hybrid (neuro-symbolic) systems
Financial & Banking	Fraud detection, credit scoring	Generative AI, symbolic AI

3.2 Aligning Business Objectives and Technology Choices

Developing AI internally starts with a clear understanding of business objectives. Organisations must define the problem they aim to solve, the expected outcomes, and operational constraints before choosing any technology. Different AI approaches serve different functions. Structured prediction tasks benefit from classical machine learning or deep learning; creative or text-generation tasks require generative models; autonomous decision-making may require agentic AI or hybrid architectures that combine symbolic and statistical methods.

The technology choice must reflect the nature of the task, data availability, accuracy requirements, integration depth, and regulatory implications. This alignment reduces the risk of selecting overly complex or inappropriate technologies and ensures that the resulting system fits the organisation's operational environment and compliance expectations.

3.3 Organisational Capabilities and Readiness

Building AI requires a set of organisational capabilities that extend beyond data science expertise. Companies must have robust data governance to ensure the quality, representativeness, and legality of the data used for training and validation. They require infrastructure to manage data pipelines, compute resources, deployment environments, and monitoring tools. Development teams must possess knowledge of security, software engineering, and MLOps practices to ensure stable and maintainable systems.

Equally important are legal, compliance, and risk management competencies. AI development requires continuous coordination between technical teams, legal counsel, and business owners to ensure that system design aligns with regulatory obligations and internal policies. Without these capabilities, organisations risk building systems that perform technically but do not meet safety, documentation, or governance requirements.

3.4 Internal Risk Types and Their Relevance

AI systems introduce new types of risks that organisations must systematically identify and manage. Understanding these risks is essential because they influence architectural choices, documentation requirements, oversight structures, and the need for mitigation measures.

Cybersecurity risks arise when AI systems expose new attack surfaces, such as model manipulation, data extraction, or system misuse. Without strong security controls, AI components can become entry points for data breaches or operational disruption. Legal and compliance risks stem from inadequate documentation, unclear explanations, or insufficient oversight, which can lead to violations of the EU AI Act or sector-specific regulations. Product liability and safety risks are particularly relevant in manufacturing, healthcare, transportation, or finance, where AI decisions may directly affect physical systems or critical outcomes.

Reliability and robustness risks occur when models perform inconsistently across environments, degrade over time, or fail in edge cases not covered during training. These issues highlight the need for proper validation, monitoring, and fallback mechanisms. Transparency and human oversight risks arise when operators do not understand or properly review AI outputs, leading to automation bias or misuse. Without structured oversight procedures, seemingly small errors can accumulate into systemic failures.

These risk types demonstrate why companies must treat AI development as a lifecycle process that integrates technical design with governance, documentation, and human supervision.

3.5 Governance, Monitoring, and Documentation

Companies building AI must implement governance and monitoring structures that ensure reliability, safety, and accountability throughout the system lifecycle. Effective governance requires clear responsibilities, human oversight protocols, and escalation paths. It integrates legal, compliance, technical, and operational roles to provide coordinated oversight of model development and deployment.

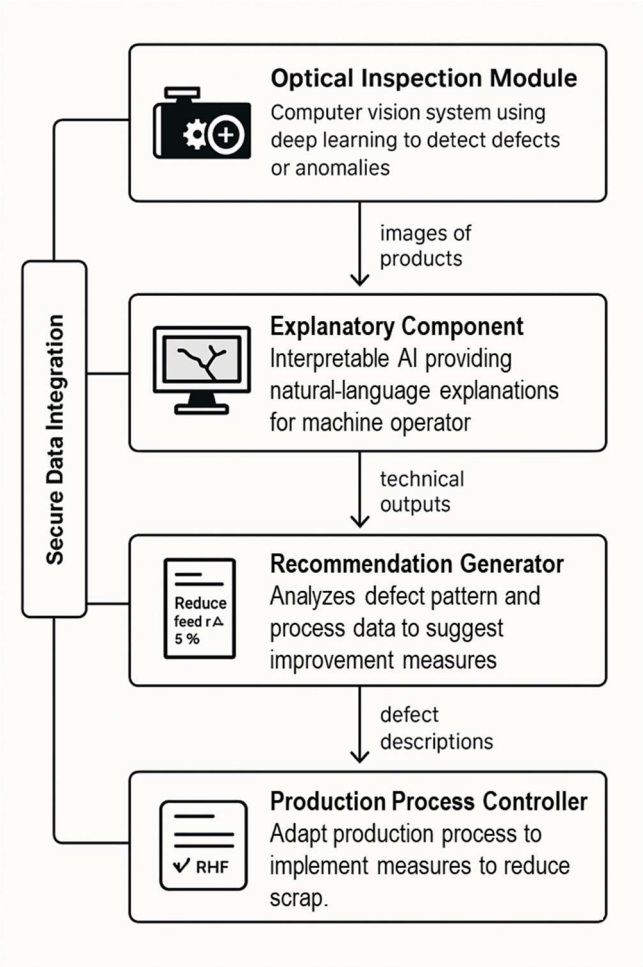
A core requirement is a Risk Management System (RMS) covering all AI risks. This should be closely coupled with the Quality Management System (QMS) in which training, testing, validation, deployment, and post-market monitoring has to be reflected. This gives input and data for the risk assessment, incident reporting, change management, and continuous improvement. Monitoring mechanisms must track model performance, operational drift, anomalies, and data quality changes. Human oversight must be formally defined to prevent overreliance on automated outputs.

Documentation plays a central role in enabling governance, transparency, and regulatory compliance. It must cover data sources, model design, training procedures, testing results, performance metrics, and known limitations. It also must include operational procedures, monitoring plans, and oversight responsibilities. Without systematic documentation, companies cannot demonstrate compliance, support audits, or maintain system reliability over time.

3.6 Example: Integrated AI Quality Assurance System

The following example reflects architectures and risk profiles already observed in industrial deployments discussed within the Charter of Trust community. An AI-based quality assurance system illustrates how internal capabilities and risks converge. Such an AI system may combine computer vision for defect detection, an interpretability module for operator understanding, and a recommendation component that proposes corrective actions. Each module requires high-quality labeled data, domain-specific fine-tuning, clear documentation, and validation across diverse operating conditions.

The AI system's complexity introduces multiple risks. Inaccurate detection may lead to scrap or safety issues; poor explanations can mislead operators; generative recommendations may be incorrect or unsafe; integration into production networks introduces cybersecurity concerns. Under the EU AI Act, such a system may be classified as high-risk, requiring robust documentation, post-market monitoring, and human oversight mechanisms, if it controls actively the production process, giving control commands to machines and intra logistics. As such a setting could have safety impacts on human workers. To sketch the consequences and efforts to run a high risk AI system, we will make this assumption. Thus, this example shows how internal capabilities must align with regulatory expectations to enable safe and effective deployment.



System Architecture and Design

The envisioned quality assurance system consists of four coordinated components:

1. **Optical Inspection Module** – A computer vision system powered by deep learning (typically convolutional neural networks, or CNNs) that detects defects or anomalies in products. It captures images via high-resolution cameras and classifies parts as conforming or non-conforming.
2. **Explanatory Component** – An interpretable AI layer (e.g., saliency maps, attention heatmaps, or a smaller language model) that translates the technical outputs of the vision model into natural-language explanations for the machine operator — describing *what went wrong* and *where* the defect occurred.
3. **Recommendation Generator** – A generative or reinforcement-learning-based subsystem that analyzes defect patterns, contextual process data (temperature, speed, material batch, etc.), and historical corrections to propose actionable improvement measures for reducing scrap and rework.
4. **Production Process Controller** – This is the component that make the system critical. If the recommendation generator consults a human team that implements the measures, the overall system will be non-critical. If the measures are translated in an automatic

reconfiguration system with all adaptations on the production line, intra logistics, e.g. the overarching production process with impact on safety and security, it will be considered a high-risk process.

These modules are orchestrated through a secure data integration layer that connects the AI components to production systems (MES/SCADA), ensuring traceability and compliance with quality standards.

3.6.1 Model Training and Data Requirements

- **Optical Inspection Model:**
  - Trained on large, labeled image datasets of both good and defective products. Data must cover variations in lighting, angles, and defect types. Augmentation techniques improve robustness. A portion of the dataset should be held out for validation to detect overfitting.

  → Output: defect detection with confidence scores and defect localization.
- **Explanatory Component:**
  - Uses interpretable AI methods and possibly a fine-tuned language model trained on production documentation, error catalogs, and maintenance reports. The model learns to map technical defect data into operator-friendly explanations.

  → Output: plain-language defect descriptions such as “Crack near weld seam due to insufficient cooling time.”
- **Recommendation Generator:**
  - Fine-tuned on historical process data, quality logs, and corrective actions. Techniques may include reinforcement learning from human feedback (RLHF) or retrieval-augmented generation (RAG) using a domain-specific knowledge base.

  → Output: improvement suggestions like “Reduce feed rate by 5%” or “Check alignment of nozzle 3.”
- **Production Process Controller:**
  - This component translates the output of the recommendation character in executable commands to the production facility. Therefore, it must have been trained on known production sequences and the interdependencies of the single production steps and their requirements.

  → Output: executable commands for production lines and intra logistics.

Each model is developed within a Quality and Risk Management System (QMS, RMS) under the EU AI Act, ensuring full documentation, version control, and post-market monitoring.

3.6.2 Operational and Compliance Risks

Such an integrated system introduces a range of operational and regulatory risks. A central challenge concerns data quality and representativeness: poorly labeled or biased image data can lead to systematic misclassification, for example by consistently underdetecting defects in certain product variants. Effective mitigation requires rigorous data validation, diverse data collection, and periodic retraining of the models.

Explainability and human oversight also present significant risks. The explanatory component may oversimplify or misinterpret technical outputs, potentially guiding operators toward incorrect conclusions. These risks can be reduced through human-in-the-loop validation, transparent user interface design, and targeted operator training.

Component	AI Act Category / Risk Level	Key Compliance Requirements	Required Documentation & Control Mechanisms
Production Process Controller	High-risk AI system under Annex III as it has direct impact on an work environment (product safety and manufacturing)	Quality and Risk Management System under Article 1 (protocol safety and validation)	• Model design and training documentation • Data provenance and labeling records • Validation results, accuracy metrics, confusion matrices • Risk log for potential misclassifications
Recommendation Generator	Decision-support component for a high risk system	Accuracy, reliability, and bias mitigation Continuous post-market monitoring Incident reporting Ethical and safe validation of generated outputs	• System architecture diagrams • Change-management documentation • Access logs • Regular security assessments
Explanatory Component	Supportive component influencing human understanding	Transparency obligations under Article 13 Human oversight controls under Art. 14 Mitigation of automation bias and misinterpretation	Documentation of interpretability methods (e.g., saliency maps, attention mechanisms) • Human–machine interface design specifications • Operator training materials • Version-controlled records of explanatory outputs
Optical Inspection (Computer Vision)	Non-high risk component for a high risk system	Accuracy, reliability, and explainability. Continuous post-market monitoring	• Model design and training documentation • Data provenance and labeling records • Validation results, accuracy metrics, confusion matrices • Risk log for potential misclassifications

3.7 Internal Control Requirements

To manage these risks, organisations must implement a comprehensive internal control framework that integrates governance structures, risk management processes, data governance, model validation, security measures, and clear human-oversight procedures. Such controls ensure that AI systems meet regulatory obligations and internal quality standards throughout their lifecycle. Data governance must guarantee accuracy, completeness, and representativeness; model validation must confirm performance under normal and adverse conditions; monitoring must detect drift and anomalies; and human oversight must remain effective through training, meaningful review mechanisms, and transparent interfaces. Post-market monitoring is essential to capture incidents, identify degradation, and trigger corrective actions.

Applied to the example system, several concrete risk areas must be addressed. A transparency gap may arise if explanations or recommendations are misleading or oversimplified, requiring operator's acknowledgment and interpretability of the audit trails. Bias in data or model outputs can result in inconsistent defect detection across variants, mitigated through diverse datasets and

fairness testing before and after deployment. Weak human oversight may lead to overreliance on AI-generated outputs, necessitating confirmation workflows, regular training, and interfaces that promote critical evaluation. Finally, insufficient post-market monitoring may allow performance drift to go unnoticed; automated dashboards, scheduled recalibration, and QMS-linked incident reporting are therefore essential. Collectively, these measures ensure that the system remains reliable, safe, and compliant in real-world operation.

3.8 Conclusions

While building AI internally offers greater control, many organisations underestimate the operational burden this entails. Internal development does not reduce regulatory obligations — it shifts them entirely into the organisation's own accountability. Companies must understand their build-versus-buy options, align technology choices with business objectives, and establish robust data governance, validation, monitoring, and documentation processes. Effective management of internal risks and controls is essential to comply with regulatory requirements and to ensure that AI systems operate reliably and safely. Organisations that develop these capabilities can build AI solutions that not only perform well but also support long-term operational and strategic goals.

4. Best Practices

Organisations that build or deploy AI systems must integrate technical, organisational, and governance measures to ensure safety, reliability, and regulatory compliance. Best practices serve as a practical framework that helps companies move from isolated AI experiments toward responsible and scalable adoption. This chapter outlines the essential best practices that enable secure, trustworthy, and resilient AI systems throughout their lifecycle.

4.1 Establishing Organisational Readiness

Effective AI adoption begins with an assessment of whether the organisation possesses the necessary capabilities, structures, and cultural foundation. Readiness includes three dimensions.

First, technical readiness concerns the ability to manage data pipelines, computing infrastructure, and software engineering practices that support model development and monitoring. Second, competency readiness reflects the availability of personnel with expertise in AI engineering, data management, cybersecurity, and domain-specific knowledge. Third, cultural readiness ensures that teams understand the opportunities and risks of AI and recognise the importance of responsible practices. Without these elements, organisations risk initiating AI projects that cannot be sustained or governed effectively.

4.2 A Phased Approach to AI Implementation

Organisational readiness for AI is less a question of ambition than of maturity. Many failures attributed to “AI risks” are, in fact, symptoms of unresolved organisational, process, or governance deficits.The first phase focuses on strategy development and governance. Organisations define their objectives, identify priority use cases, establish oversight structures, and launch AI literacy programmes. This phase also includes the development of risk management procedures aligned with regulatory obligations.

The second phase centers on pilot projects. Selected use cases are implemented in controlled environments, supported by monitoring and evaluation mechanisms. These pilots allow teams to validate performance, understand integration challenges, and refine governance processes before broader deployment.

Once the organisation has established confidence and proven value, the final phase involves scaling. Systems are deployed across additional units or processes, supported by expanded infrastructure, updated governance mechanisms, and continuous monitoring. A phased approach ensures that technical and organisational maturity grow in parallel.

4.3 Governance and Cross-Functional Collaboration

AI systems require governance frameworks that define responsibilities, decision-making processes, and oversight mechanisms. Effective governance integrates technical, legal, compliance, risk management, and business functions to ensure that AI development aligns with organisational goals and regulatory requirements.

Central elements of governance include an AI oversight committee responsible for strategic direction, risk tolerance, and escalation procedures. Operational governance teams handle daily compliance tasks, documentation, monitoring, and incident reporting. Regular coordination between teams ensures that risks are identified early, and that decisions are informed by diverse perspectives. Clear lines of accountability support consistent and predictable governance across the

organisation.

4.4 Responsible AI Principles

Organisations must anchor their AI development in responsible AI principles that guide decisions throughout the lifecycle. These principles should be translated into concrete processes, tools, and controls.

Fairness requires assessing and mitigating bias in data and model outputs. Transparency requires that systems provide meaningful explanations, documentation, and communication of capabilities and limitations. Accountability requires assigning clear responsibility for model design, deployment, and oversight. Privacy and security demand technical safeguards to protect data and prevent unauthorised access. Safety requires validation of system outputs and prevention of harmful or unintended behaviour. Embedding these principles into practice builds trust and supports compliance with regulation and ethical expectations.

4.5 Risk Management and Quality Assurance

AI-specific risks must be identified, assessed, and mitigated throughout the system lifecycle. A comprehensive risk management framework addresses technical performance risks, ethical risks, regulatory risks, and operational vulnerabilities.

Quality assurance processes ensure that data is accurate and representative, models are validated under realistic and edge-case conditions, and performance metrics are regularly monitored. Testing must include robustness checks, stress scenarios, integration testing, and assessments of system behaviour under drift. Incident response procedures support quick identification and correction of failures. Quality management systems such as ISO 9001 or ISO/IEC 42001 provide helpful structures for establishing consistent and auditable processes.

4.6 Secure and Compliant AI Deployment

Deployment of AI systems introduces new risks related to integration, security, and operational stability. Organisations should adopt deployment architectures that minimise exposure of sensitive data and ensure reliable performance. This includes secure data pipelines, encrypted communication, controlled access to models and APIs, and audit-ready logging of system events.

Operational environments must support monitoring, anomaly detection, and rollback mechanisms. Deployment processes should be reviewed for compliance with cybersecurity frameworks and sector-specific regulations. For high-risk systems, traceability of all model versions, training data, and changes must be ensured. Alignment between production environments, monitoring systems, and governance structures is essential for maintaining safety and compliance.

4.7 General-Purpose and Agentic AI Systems

Organisations using general-purpose AI (GPAI) or agentic AI systems must implement additional checks due to their high capability and unpredictability. GPAI models require documentation of training data provenance, performance limitations, and known risks. They must be assessed for systemic impacts and integrated into governance frameworks.

Agentic AI systems, which can autonomously execute tasks, introduce specific risks such as loss of oversight, cascading failures, or unintended tool usage. Organisations that deploy agentic AI without redefining oversight are effectively outsourcing risk decisions to system behaviour. These systems require testing for instruction handling, memory persistence, tool access boundaries, and

misalignment scenarios. Continuous red teaming helps identify vulnerabilities in autonomy, reasoning, and tool execution. Controls must ensure that agents operate within defined boundaries and that human oversight remains effective.

4.8 Continuous Improvement and Monitoring

AI systems evolve over time as data distributions, user interactions, and business environments change. Continuous monitoring is essential to detect model drift, performance degradation, or emerging risks. Organisations should establish dashboards and alerts to support real-time oversight.

Periodic reviews ensure that documentation, risk assessments, and governance procedures remain up to date. Lessons learned from audits, incidents, and user feedback should be incorporated into ongoing improvements. Regulatory changes and new technical standards must be monitored and integrated into processes. Continuous improvement ensures that AI systems remain effective, compliant, and aligned with organisational objectives.

4.9 Proactive Supply Chain Risk Management

Organizations should establish robust processes to identify, assess, and mitigate supply chain risks associated with AI systems. This includes:

a) Diversification of Suppliers: Avoid over-reliance on single vendors for critical hardware, software, or cloud services.

b) Continuous Monitoring: Track regulatory changes and geopolitical developments that may affect sourcing or technology access.

c) Supplier Audits and Due Diligence: Regularly evaluate suppliers for compliance with legal, ethical, and security standards.

d) Contingency Planning: Develop strategies for rapid replacement of restricted or unavailable components, including alternative sourcing and modular system design.

4.10 Culture, Training, and Change Management

Sustainable AI adoption depends on a workforce that understands AI systems, their limitations, and the organisation's responsibilities. Training programmes should be tailored to different roles, from developers and operators to managers and executives. Training should cover AI fundamentals, responsible practices, risk awareness, and compliance obligations.

Cultural development includes fostering open communication about risks, encouraging reporting of concerns, and recognising responsible behaviour. Change management processes help integrate AI into existing workflows, address resistance, and embed responsible AI values into daily operations.

4.11 Conclusions

Best practices for AI encompass readiness, governance, risk management, responsible design, secure deployment, and continuous improvement. Organisations that integrate these practices into their operations are better equipped to build AI systems that are safe, compliant, effective, and aligned with long-term strategic goals. Responsible AI is not a single activity but an ongoing discipline that ensures trustworthy and sustainable innovation.

5. Recommendations

Organisations must approach regulatory compliance, such as the EU AI Act requirements, as a strategic transformation rather than a regulatory checklist. The AI Act fundamentally reshapes how AI systems must be designed, documented, governed, and monitored. To meet these expectations and realise competitive benefits, organisations should adopt a phased set of recommendations addressing immediate, medium-term, and long-term priorities.

5.1 Immediate Priorities

The EU AI Act follows a phased implementation timeline, and several obligations take effect well before the full application in 2027. Early preparation is essential. Organisations should begin by conducting a complete inventory of all AI systems—internal, third-party, embedded, or experimental—and classify them by risk category, ideally embedding them in existing inventories. This provides the foundation for all subsequent compliance planning.

A cross-functional compliance team should be established to ensure coordinated implementation across legal, technical, risk, and business functions. AI literacy training must also begin immediately for all personnel involved in developing, deploying, or supervising AI systems; this requirement is already in force. Finally, organisations must confirm that none of their systems fall under the Act's list of prohibited practices, as violations carry the highest penalties.

5.2 Strengthening Risk Management and Governance

Robust governance is central to sustainable compliance. Organisations should adopt a multidimensional risk assessment methodology that evaluates performance risks, ethical implications, regulatory exposure, and operational continuity. This assessment must be iterative and integrated into every stage of the AI lifecycle.

Governance structures should include an executive-level AI committee responsible for strategic decisions and risk tolerance, supported by operational teams managing day-to-day compliance, documentation, and incident response. Clear coordination mechanisms must align legal, technical, and business stakeholders, ensuring consistent oversight and rapid decision-making.

5.3 Implementing Quality and Documentation Frameworks

A comprehensive Quality Management System (QMS) tailored to AI is essential. Organisations should define procedures for data governance, model validation, bias detection, monitoring, and corrective actions. Integration with existing quality frameworks—such as ISO 9001 or sector-specific standards—avoids duplication and strengthens organisational maturity. Adoption of ISO/IEC 42001 can provide a structured foundation for an AI management system.

Documentation plays a critical role in demonstrating compliance. Automated evidence management systems should track model versions, testing results, monitoring outputs, and governance decisions across the lifecycle. Documentation must remain audit-ready, current, and accessible to authorised stakeholders.

5.4 Managing Financial Exposure and Building Competitive Advantage

The AI Act introduces substantial penalties—up to €35 million or 7% of global turnover—making proactive compliance the most cost-effective strategy. Organisations should balance investment in compliance with the financial and reputational risks of non-compliance. SMEs should leverage

simplified procedures and national support programmes to reduce administrative burden.

Early compliance can also create competitive advantages. Organisations that establish strong governance and transparency practices can differentiate their AI products, accelerate innovation through reduced regulatory uncertainty, and build trust with customers, partners, and regulators.

5.5 Ensuring Long-Term Resilience and Future-Proofing

Because AI regulation and technology will continue evolving, compliance strategies must remain adaptive. Organisations should design flexible policies, modular technical architectures, and scalable governance processes that can adjust to updated requirements. Continuous regulatory monitoring and participation in standards-setting initiatives support early awareness and influence future developments.

Building internal capabilities is equally important. A sustainable compliance culture requires strong leadership commitment, empowered employees, and role-specific training programmes. Encouraging ethical awareness and responsible AI behaviour throughout the organisation ensures that compliance becomes an integrated practice rather than an isolated project.

5.6 Roadmap for Implementation and Indicators of Governance Maturity

A phased approach supports structured and timely compliance:

- **Phase 1 (0–6 months):** Build foundations through system inventory, governance setup, AI literacy programmes, and elimination of prohibited practices.
- **Phase 2 (6–18 months):** Develop core capabilities, including risk management frameworks, QMS processes, documentation systems, monitoring mechanisms, and preparation for high-risk system compliance.
- **Phase 3 (18–36 months):** Achieve full compliance for high-risk systems, refine processes based on operational experience, and leverage compliance as a source of competitive advantage.
- Success depends on measurable indicators. Organisations should track their adherence to AI Act requirements, the effectiveness of risk mitigation, operational performance, and cultural adoption of responsible AI practices. Monitoring these indicators ensures continuous improvement and readiness for regulatory oversight.

5.7 Conclusions

Preparing for upcoming AI legislation, such as the EU AI Act, requires immediate action, structured governance, and long-term organisational commitment. Companies that adopt proactive, comprehensive compliance strategies will not only reduce regulatory risk but also strengthen their competitive position in an increasingly accountability-driven AI landscape. Those who delay face escalating penalties, operational uncertainty, and reduced market trust. By treating compliance as an enabler of responsible innovation, organisations can shape AI systems that are safe, transparent, and aligned with societal values.