Integrity360

your **security** in mind

Welcome

AI Risk Management in practice: From Governance to Testing

Our speakers

Rich Ford

Chief Technology Officer

Integrity360

Eugenio Bonzi

Cyber Risk & Assurance Advisor

Integrity360

Alexandros Koutroutsios

Penetration Test Team Leader

Integrity360

AI is on the rise!

AI's expanding influence is reshaping industries and everyday life, clearly improving efficiency and innovation. Its increasing adoption across all sectors reflects a strong trend towards intelligent automation and data-driven solutions. This growth underlines AI's transformative potential, subject to evolving regulatory frameworks and ethical considerations, promising significant advances in productivity and user experience.

Everyday

- Virtual assistant
- Recommendation systems
- Automatic translation
- Navigation
- Facial recognition
- Photo filters and effects

Business

- Process automation
- Customer service
- Data analysis
- Personalized marketing
- Fraud detection
- Supply chain management

Healthcare

- Medical diagnoses
- Personalized medicine
- Robotic surgery
- Remote patient monitioring

Education

- Personalized learning
- Intelligent tutoring systems
- Automatic evaluation
- Accessibility

Cyber security

- Threat detection and prevention
- Vulnerability management
- Incident response
- Identity and Access Management (IAM)
- Security Operation (SecOps)
- Cloud security
- Data Security

Integrity360
your security in mind

AI Risk Management in practice: From Governance to Testing

A comprehensive approach to managing artificial intelligence risks across your enterprise

Approach

1

The AI risk management challenge

Understanding the unique risk factors of enterprise AI adoption and the critical need for integrated governance approaches.

2

Governance frameworks

Examining NIST AI RMF, ISO/IEC 42001 standard and EU AI Act regulation and their practical application to AI systems.

3

Technical validation

Exploring penetration testing methodologies and security controls specific to AI applications.

4

Implementation roadmap

Establishing organizational ownership and developing a repeatable model for AI risk management.

AI is on the rise

The AI risk management challenge

The accelerated adoption of AI technologies introduces unique risk vectors that traditional security approaches may not adequately address:

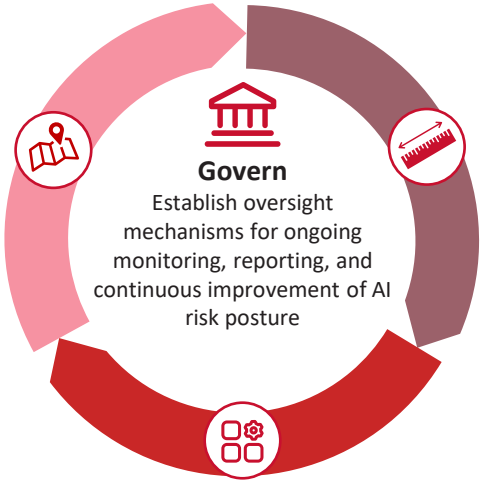
- **Data privacy** concerns with training datasets
- **Model bias** and decision transparency issues
- **Prompt injection** vulnerabilities
- **Third-party API** dependencies
- **Regulatory compliance** challenges across jurisdictions

Organizations must develop integrated approaches that combine strategic oversight with technical validation to ensure AI deployments remain secure and compliant.

Integrity360
your security in mind

NIST AI RMF

Map
Identify context, capabilities, limitations, and potential impacts of AI systems within your enterprise landscape.



Measure
Quantify AI risks through assessment methodologies tailored to specific use cases and organizational contexts.

Manage
Implement governance structures, policies, and technical controls to mitigate identified risks.

The NIST AI risk management framework provides a structured approach to identifying, assessing, and managing risks associated with AI systems throughout their lifecycle.

Governance framework

ISO/IEC 42001

ISO/IEC 42001 establishes requirements for AI management systems, enabling organizations to:

- **Demonstrate** responsible AI governance
- **Build trust** with stakeholders and regulators
- **Enhance** AI system quality and reliability
- **Mitigate** potential risks and harms

This standard employs the Plan-Do-Check-Act cycle for continuous improvement and aligns with existing management system standards like ISO 27001 for information security.

Integrity360
your security in mind

EU AI act

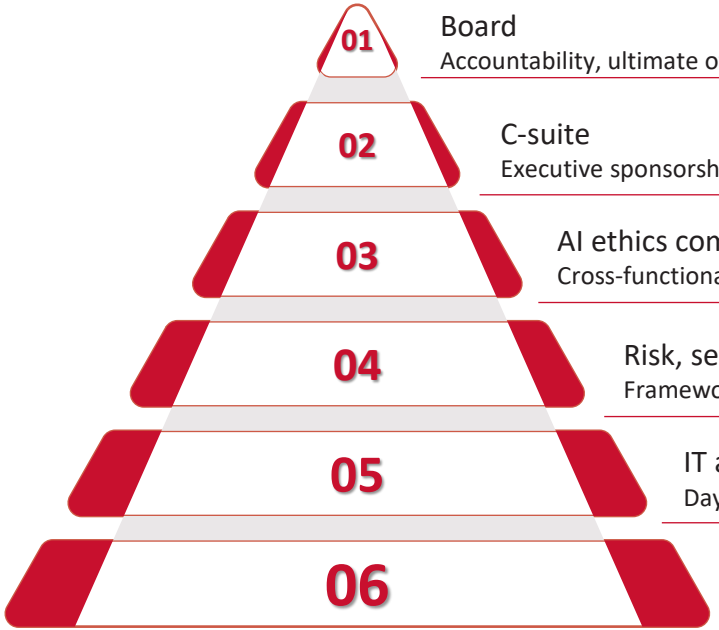
The European union AI act is the world's first comprehensive legal framework for AI, aiming to ensure AI systems are safe, transparent and human-centric.
It adopts a risk-based approach, categorizing AI systems into four levels:

- **Unacceptable** risk (e.g. manipulative AI)
- **High-risk** (must be regulated)
- **Limited risk** (subject to lighter transparency obligations)
- **Minimal risk** (unregulated, including the majority of AI applications available on the EU market)

High-risk AI systems, such as those used in critical infrastructure or law enforcement, face stringent obligations including robust risk management, data governance, human oversight, and cybersecurity measures.

Organisational risk

Ownership of AI risk



Board
Accountability, ultimate oversight and strategic direction

C-suite
Executive sponsorship and resource allocation

AI ethics committe
Cross-functional governance and policy development

Risk, security and compliance
Framework implementation and technical controls

IT and business units
Day-to-day implementation and operational compliance

End users
Responsible use

Real-world application

Microsoft Copilot

Applying governance frameworks to Microsoft Copilot deployment requires balancing productivity benefits with appropriate risk controls.

1

Risk assesment

- Data leakage potential through prompts
- Intellectual property exposure
- Regulatory compliance implications
- Integration with sensitive systems

2

Governance control

- Microsoft Purview integration
- Data Loss Prevention policies
- Role-based access controls
- Tenant-level configuration options

3

Implementation steps

- Pilot deployment with controlled user groups
- Documented approval workflows
- User training on acceptable use
- Monitoring and audit procedures

Technical validation

Ai-specific penetration testing

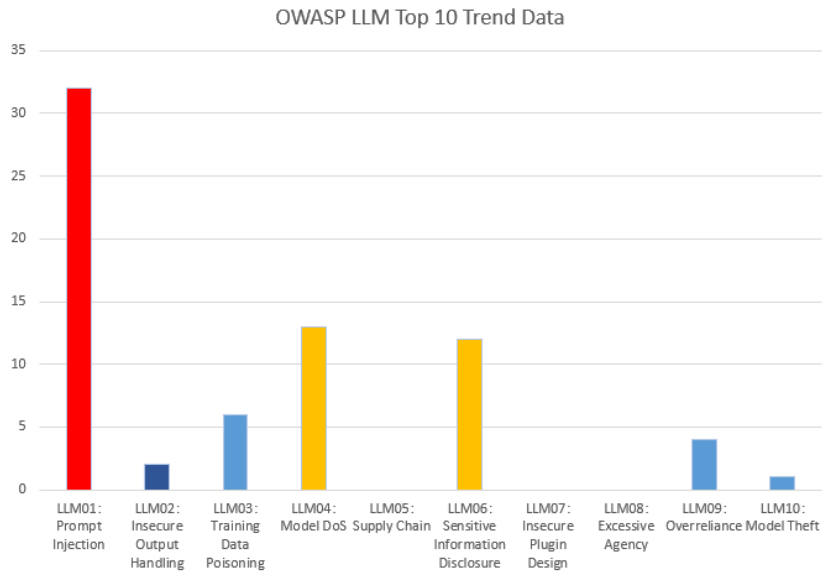
AI-specific penetration testing requires specialized methodologies that go beyond traditional application security testing to address unique AI vulnerabilities.

Unique Threats

- Prompt injection attacks to bypass content filters
- Data extraction attempts through indirect queries
- Model manipulation to produce harmful outputs

Business Impact

- Intellectual property theft
- Model integrity loss
- Regulatory non-compliance



Technical validation

Practical testing approach for AI systems

Testing areas

- Model and data
- Interfaces and APIs
- LLMs and GenAI

Defence/ mitigation

- Layered safety constrains
- Secure API design and model hardening
- Robust prompt filtering and output guardrails
- Third-party testing

Implementation roadmap

Integrating AI risk management model

Document AI use cases

Inventory all AI systems and categorize by risk level

Apply governance framework

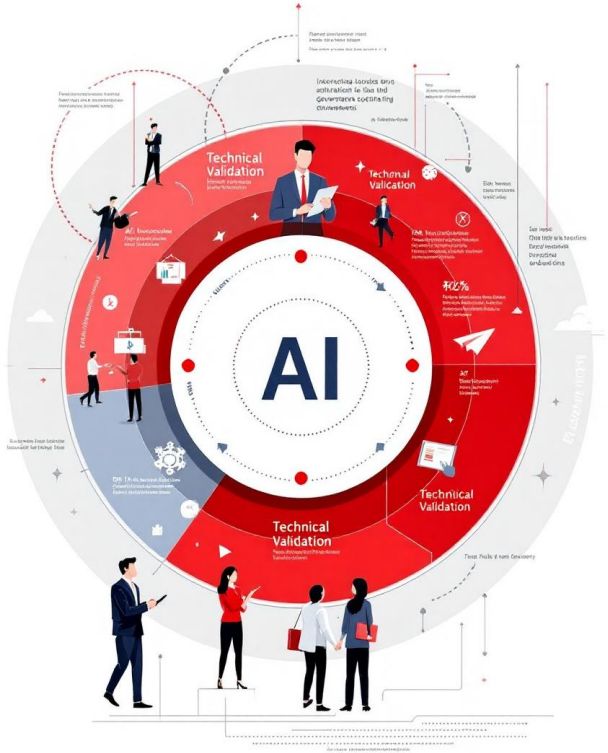
Implement NIST/ISO controls based on risk assessment

Validate technical controls

Conduct penetration testing and security assessment

Continuous monitoring

Establish metrics and review procedure for ongoing compliance



Integrity360
your security in mind

Key takeways

1

Governance & testing

Combine standards-based governance with technical validation for comprehensive AI risk management

2

Shared responsibility

Establish clear organizational ownership across technical and business functions

3

Continuous process

Implement a repeatable model that evolves with AI technologies and regulatory landscape

Next steps

- Conduct an inventory of current AI systems
- Assess your organizational AI governance maturity
- Identify high-priority risk areas for immediate action
- Develop a roadmap for implementing integrated controls

Contact us for a complimentary AI risk assessment consultation

Q&A

Upcoming webinars

Register

Integrity360
your security in mind

Thank you

Next steps

Your Integrity360 Account Manager will reach out to you for any further questions you might have.

Webinar recording and useful resources will be emailed to you

If you have any other questions, please email marketing@integrity360.com