TEMET

end-to-end IT security

AI Compliance Essentials: Standards and Emerging Regulations

Berne, 22.10.2024

Speaker

Bruno Blumenthal

Dipl.-Ing. FH in Computer Science
CISM, CISA, CISSP
Partner, Member of the Executive Board at TEMET AG

Main Topics:

- Security Strategy & Architecture
- Information Security Management Systems (ISMS)
- Cybersecurity & Security Operations

Contact:

Mobile: +41 78 859 57 15
E-Mail: bruno.blumenthal@temet.ch

Standards

ISO/IEC

Joint Technical Committee 1: Information Technology

Subcommittee 42: Artificial Intelligence

ISO/IEC

Joint Technical Committee 1: Information Technology

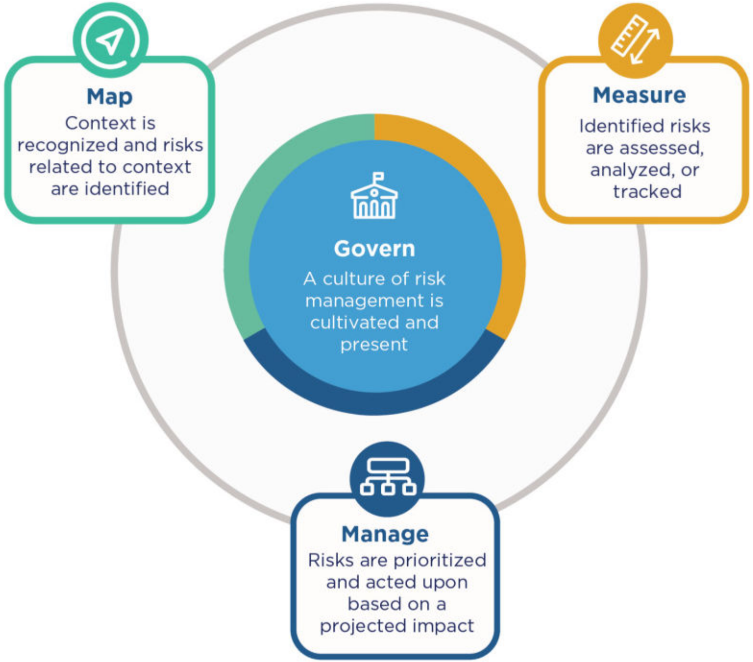
Subcommittee 42: Artificial Intelligence

General Topics
ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

Risk Management and Governance	
ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system	**ISO/IEC DIS 42005 (Draft)** Information technology — Artificial intelligence — AI system impact assessment
ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations	**ISO/IEC 23894:2023** Information technology — Artificial intelligence — Guidance on risk management

NIST AI Risk Management Framework

Advance the safe, secure, and trustworthy development and use of AI.



Map
Context is recognized and risks related to context are identified

Measure
Identified risks are assessed, analyzed, or tracked

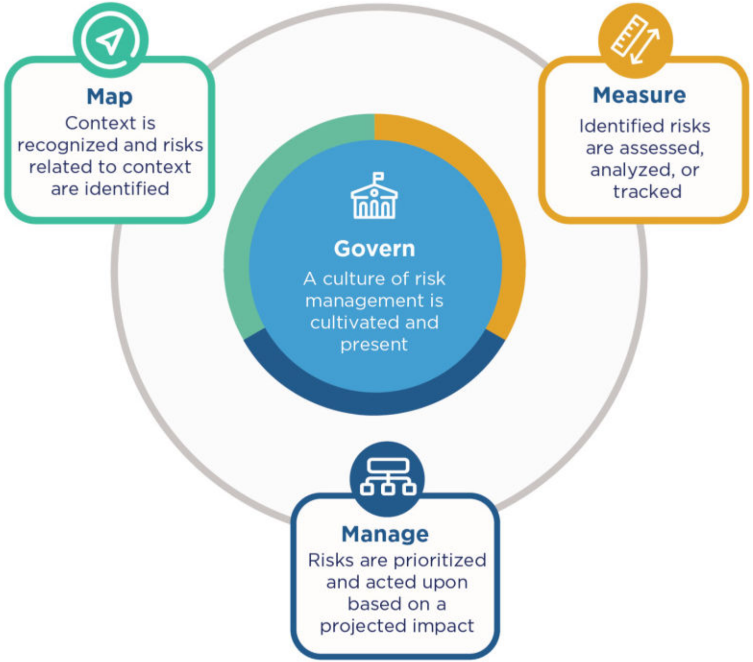
Govern
A culture of risk management is cultivated and present

Manage
Risks are prioritized and acted upon based on a projected impact

NIST AI Risk Management Framework

Advance the safe, secure, and trustworthy development and use of AI.

NIST AI 100-1: AI RMF 1.0
4 Functions, 19 Categories, 69 Subcategories



Map
Context is recognized and risks related to context are identified

Measure
Identified risks are assessed, analyzed, or tracked

Govern
A culture of risk management is cultivated and present

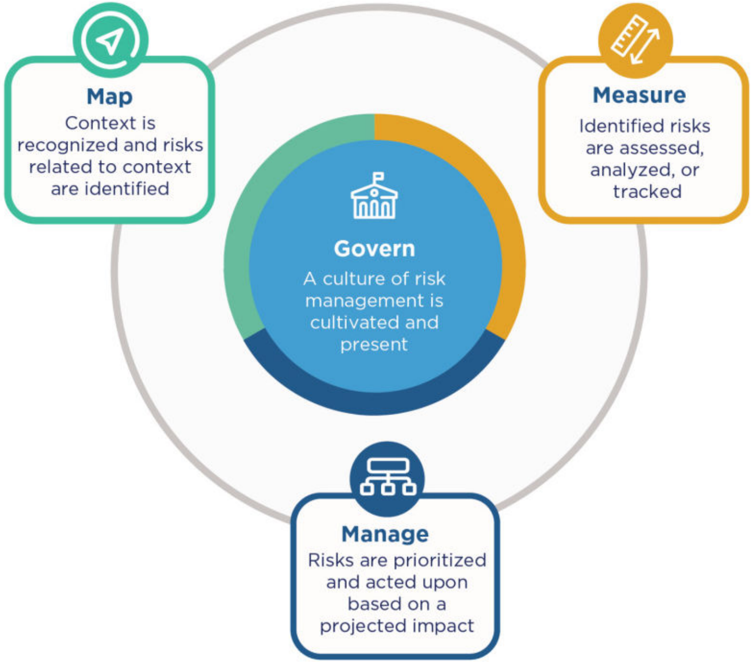
Manage
Risks are prioritized and acted upon based on a projected impact

NIST AI Risk Management Framework

Advance the safe, secure, and trustworthy development and use of AI.

NIST AI 100-1: AI RMF 1.0
4 Functions, 19 Categories, 69 Subcategories

NIST AI 600-1
Generative Artificial Intelligence Profile



Map
Context is recognized and risks related to context are identified

Measure
Identified risks are assessed, analyzed, or tracked

Govern
A culture of risk management is cultivated and present

Manage
Risks are prioritized and acted upon based on a projected impact

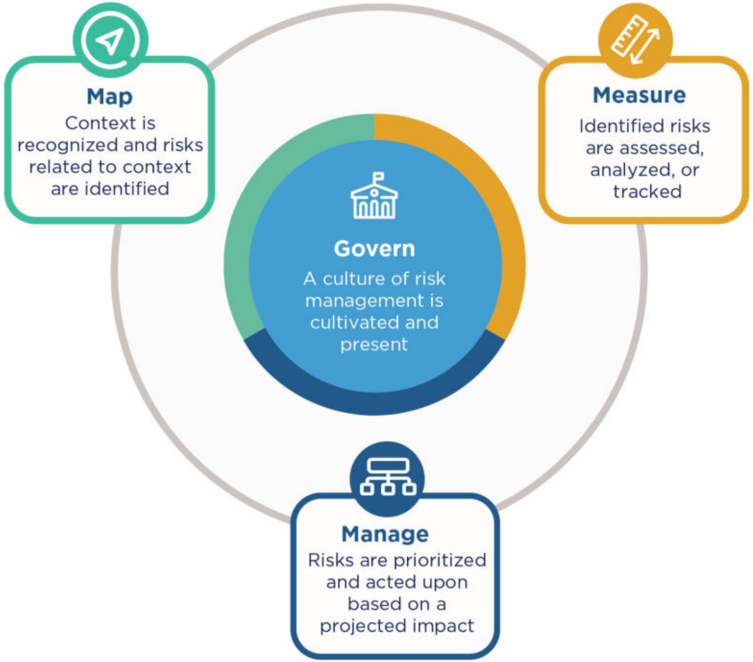
NIST AI Risk Management Framework

Advance the safe, secure, and trustworthy development and use of AI.

NIST AI 100-1: AI RMF 1.0
4 Functions, 19 Categories, 69 Subcategories

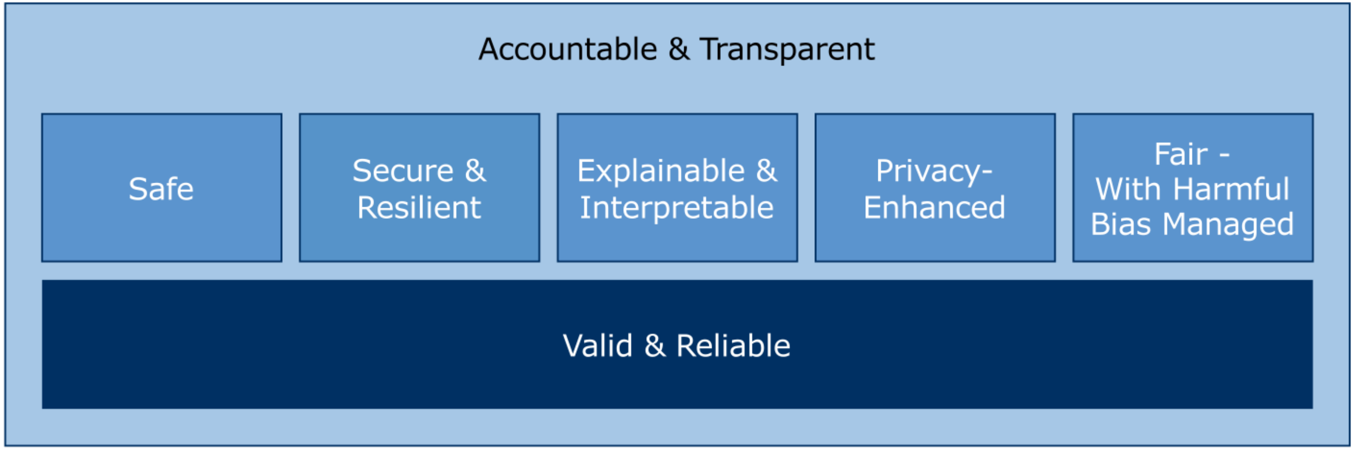
NIST AI 600-1
Generative Artificial Intelligence Profile

NIST AI RMF Playbook
Implementation guidance


Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

Trustworthy AI

Properties of Trustworthy AI



Accountable & Transparent

Safe

Secure & Resilient

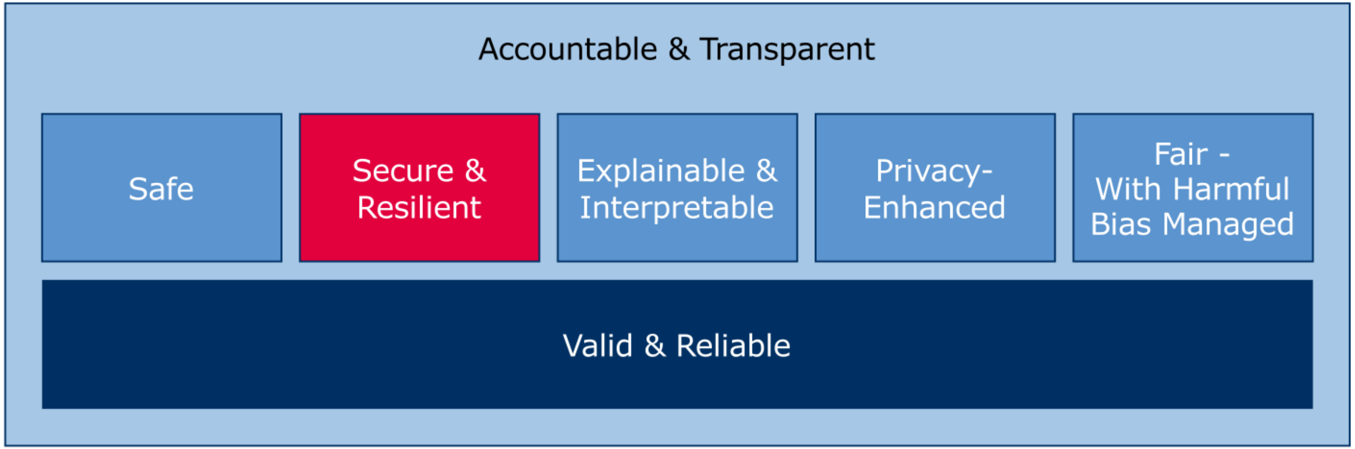
Explainable & Interpretable

Privacy-Enhanced

Fair - With Harmful Bias Managed

Valid & Reliable

Properties of Trustworthy AI



Accountable & Transparent

Safe

Secure & Resilient

Explainable & Interpretable

Privacy-Enhanced

Fair - With Harmful Bias Managed

Valid & Reliable

Regulations

Focus of the regulators

Focus of the regulators

Automated Decision Making

Systems that make decisions on behalf of humans.

- Screening resumes
- Determining eligibility for credit
- Making parole decisions
- Predictive policing
- Predictive maintenance

Focus of the regulators

Automated Decision Making

Systems that make decisions on behalf of humans.

- Screening resumes
- Determining eligibility for credit
- Making parole decisions
- Predictive policing
- Predictive maintenance

Recognition Technologies

Technology to identify objects or humans.

- Face or voice recognition
- Gait analysis
- User behavior analysis

Focus of the regulators

Automated Decision Making

Systems that make decisions on behalf of humans.

- Screening resumes
- Determining eligibility for credit
- Making parole decisions
- Predictive policing
- Predictive maintenance

Recognition Technologies

Technology to identify objects or humans.

- Face or voice recognition
- Gait analysis
- User behavior analysis

Manipulative or Deceptive Use

Systems to manipulate or deceive humans or systems.

- Deep fakes
- Voice cloning
- Fake social media accounts
- Systems that impersonate humans

EU AI Act

Chapter II: Prohibited AI Practices

- Exploiting, discriminating, manipulating, or deceiving people
- Non-targeted use of biometric recognition technology
- Exceptions for law enforcement

EU AI Act

Chapter II: Prohibited AI Practices

- Exploiting, discriminating, manipulating, or deceiving people
- Non-targeted use of biometric recognition technology
- Exceptions for law enforcement

Chapter III: High-Risk AI System

- A list of high-risk applications is provided
  - Biometrics, safety, gatekeepers, law enforcement, justice, and democracy
- Third-party conformity assessment must be performed
- A risk management system must be in place
- Article 15: Accuracy, Robustness and Cybersecurity

Council of Europe

The Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law

Council of Europe

The Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law

- Created by member states and non-member states
- First-ever international legally binding treaty in this field, but not yet ratified by any member or non-member state

Council of Europe

The Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law

- Created by member states and non-member states
- First-ever international legally binding treaty in this field, but not yet ratified by any member or non-member state
- Focus on:
  - Human dignity and individual autonomy
  - Equality and non-discrimination
  - Respect for privacy and personal data protection
  - Transparency and oversight
  - Accountability and responsibility
  - Reliability
  - Safe innovation

Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI

Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI
- Four possible outcomes

Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI
- Four possible outcomes
  - Do nothing

Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI
- Four possible outcomes
  - Do nothing
  - Do something only for the government

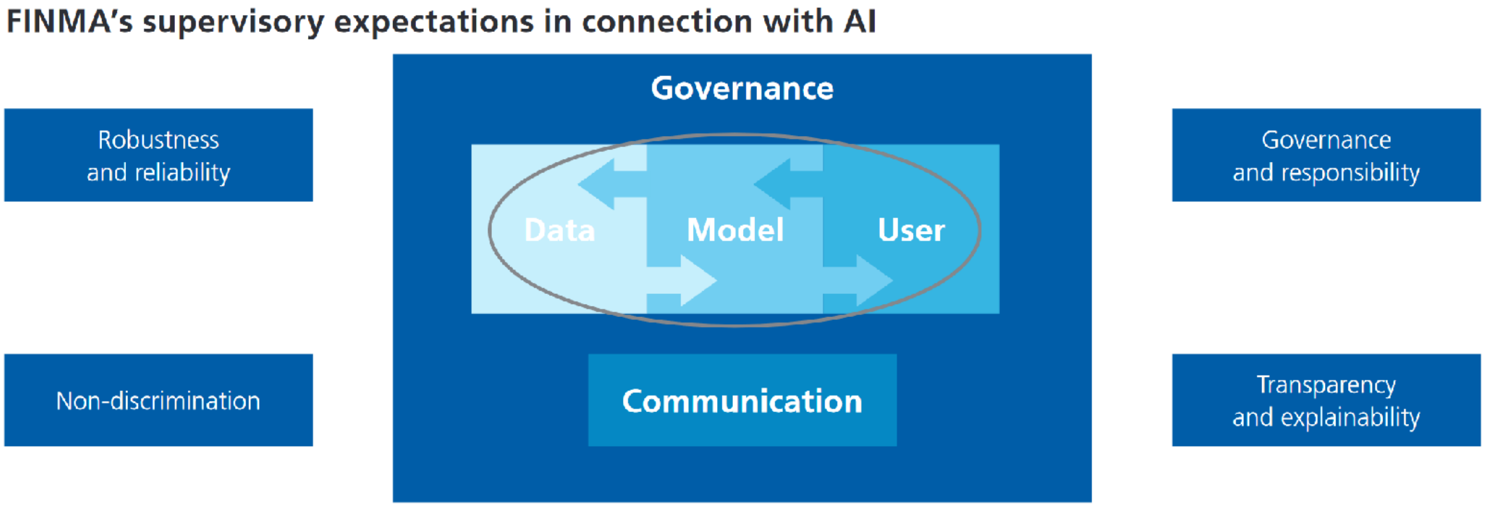
Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI
- Four possible outcomes
  - Do nothing
  - Do something only for the government
  - Create an overarching regulation

Switzerland

- Swiss government observes
- Federal Council wants to harness the potential of AI while minimizing the risks to society
- Currently preparing an overview of possible regulatory approaches to AI
- Four possible outcomes
  - Do nothing
  - Do something only for the government
  - Create an overarching regulation
  - Leave regulation to sector-specific regulators

FINMA



FINMA's supervisory expectations in connection with AI

Robustness and reliability

Non-discrimination

Governance

Data

Model

User

Communication

Governance and responsibility

Transparency and explainability

Conclusion

???