Artificial Intelligence (AI) from a Governance Perspective

SiGMA TECHNOLOGY

AI Opportunities and Challenges

The adoption of artificial intelligence (AI) around the world is transforming industries, governments, and daily life across multiple dimensions. The global adoption of AI is also accompanied by significant challenges, including ethical concerns, the need for robust data privacy protections, and the potential for job displacement due to automation. Moreover, there is a growing emphasis on creating frameworks for AI governance to ensure that AI development is ethical, secure, and beneficial to society.

From a governance perspective, two prominent frameworks stand out: The International Standard Organization (ISO) and the National Institute of Standards and Technology (NIST).

The Emerging Standards in AI from a Governance Perspective

ISO

ISO/IEC 42001 is an international standard that specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations. It is designed for entities providing or utilizing AI-based products or services, ensuring responsible development and use of AI systems.

NIST

National Institute of Standard and Technology (NIST) has introduced Artificial Intelligence Risk Management Framework (AI RMF 1.0). The goal of the AI RMF is to offer a resource to the organizations designing, developing, deploying, or using AI systems to help manage many risks of AI and promote trustworthy and responsible development and use of AI systems.

AI Governance Risks

Various AI Governance Risks include:

- Garbage in Garbage out (GIGO) –algorithm correctly performing incorrect tasks without AI strategy & purpose alignment
- Unintended model mutations due to gaps in algorithm governance across model evaluation and continuous training processes
- Ineffective development, performance and improvement challenges due to gaps in stakeholder feedback loop (upstream or downstream)
- Unreliable and unpredictable computing outputs due to improper or incomplete training processes and algorithm lifecycle management
- Lack of ethical AI governance considerations can lead to bias or other model errors impacting users and society (ethical impact assessment)
- Model availability and continuity challenges due to lack of contingency planning and recovery procedures
- Lack of privacy review considerations in data processing can lead to regulatory noncompliance (processing PII, PHI, PCI or youth)
- Lack of consideration of sector specific risks and impacts arising from the application of the AI model (i.e., unintended impacts on OH&S, Industrial Operations, etc., as applicable)

Overview of ISO/IEC 42001 AIMS

Establishing and Supporting the AIMS (Plan)

Clause 4 Context of the organization
4.1 Understanding the organization and its context
4.2 Understanding the needs and expectations of interested parties
4.3 Determining the scope of the AI management system
4.4 AI management system

ISO/IEC 38507 AI – Governance implications of the use of AI

Clause 5 Leadership
5.1 Leadership and commitment
5.2 AI Policy
5.3 Roles, responsibilities and authorities

Clause 6 Planning
6.1 Actions to address risks and opportunities
6.2 AI objectives and planning to achieve them
6.3 Planning of changes
ISO/IEC 23894 AI – Guidance on risk management

Clause 7 Support(ing)
7.1 Resources
7.2 Competence
7.3 Awareness
7.4 Communication
7.5 Documented information

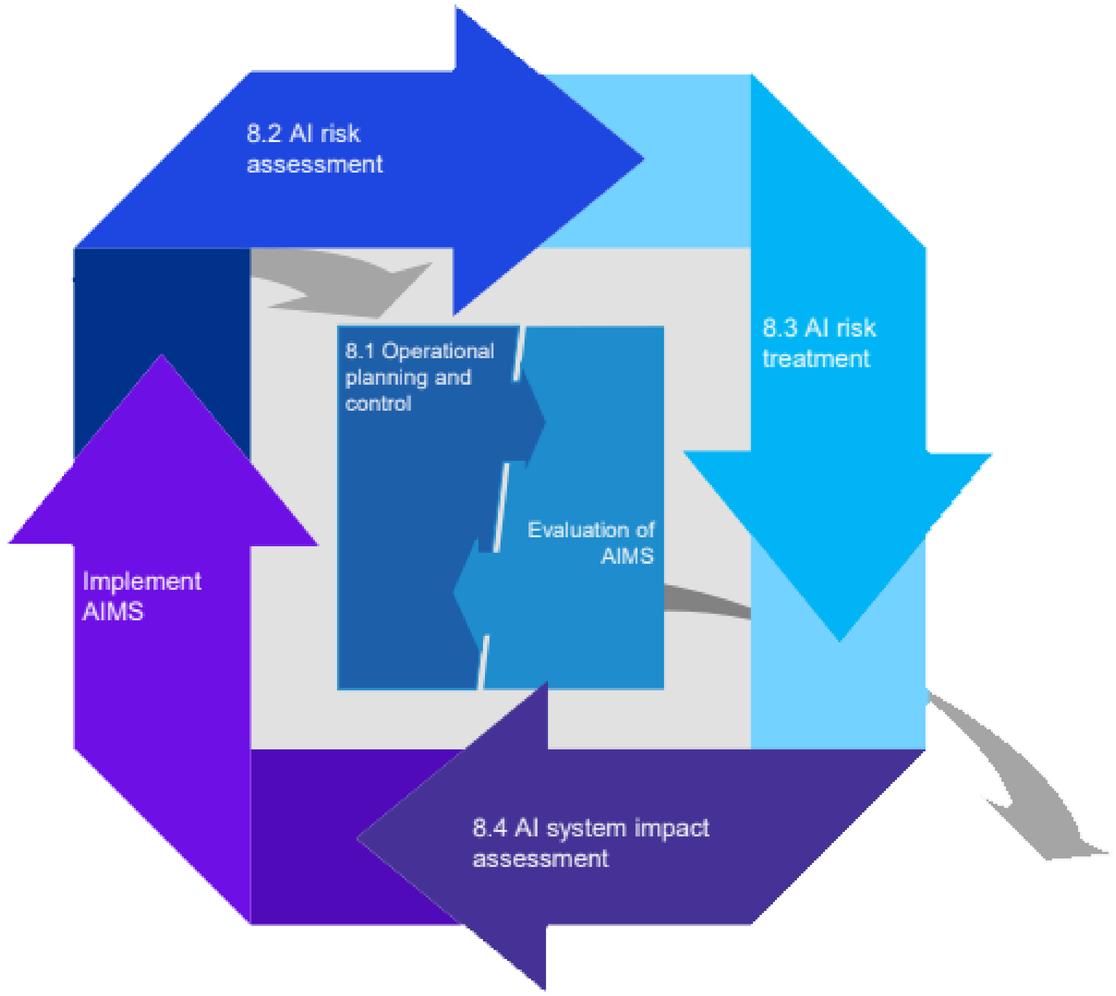
Annex A
Reference AI control objectives and controls, the controls are optional, the controls selected can come from somewhere else

Annex B
Implementation guidance for AI controls

Annex C
Potential AI-related organizational objectives and risk sources

Annex D
Use of AI management system across domains or sectors

Integration of AIMS with other management systems



Clause 9 Performance Evaluation (Check)

9.1 Monitoring, measure-ment,analysis and evaluation

9.2 Internal Audit

External Audit

Identification of Non-conformities (Resulting in a Corrective Action Plans)

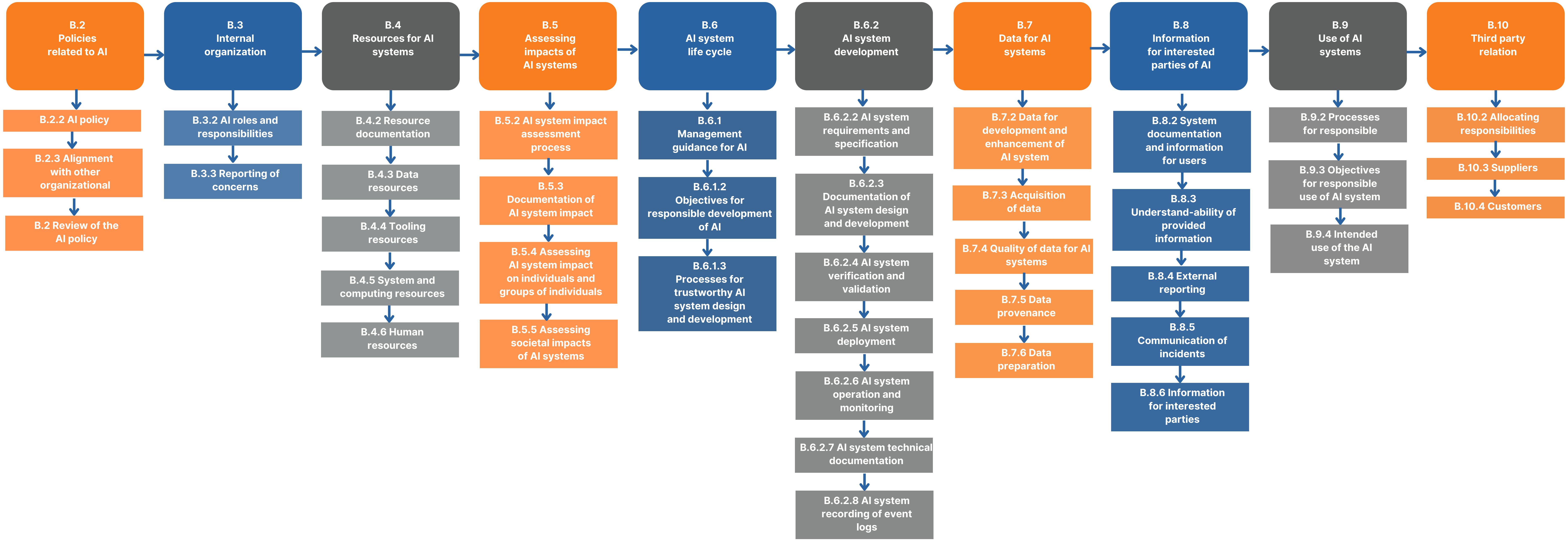
9.3 Management Review

Clause 10 Improvement (Act)

10.1 Nonconformity and corrective action
10.2 Continual improvement

SiGMA TECHNOLOGY

ISO/IEC 42001 ANNEX A: AI Control Objectives



B.2 Policies related to AI
- B.2.2 AI policy
- B.2.3 Alignment with other organizational
- B.2 Review of the AI policy

B.3 Internal organization
- B.3.2 AI roles and responsibilities
- B.3.3 Reporting of concerns

B.4 Resources for AI systems
- B.4.2 Resource documentation
- B.4.3 Data resources
- B.4.4 Tooling resources
- B.4.5 System and computing resources
- B.4.6 Human resources

B.5 Assessing impacts of AI systems
- B.5.2 AI system impact assessment process
- B.5.3 Documentation of AI system impact
- B.5.4 Assessing AI system impact on individuals and groups of individuals
- B.5.5 Assessing societal impacts of AI systems

B.6 AI system life cycle
- B.6.1 Management guidance for AI
- B.6.1.2 Objectives for responsible development of AI
- B.6.1.3 Processes for trustworthy AI system design and development

B.6.2 AI system development
- B.6.2.2 AI system requirements and specification
- B.6.2.3 Documentation of AI system design and development
- B.6.2.4 AI system verification and validation
- B.6.2.5 AI system deployment
- B.6.2.6 AI system operation and monitoring
- B.6.2.7 AI system technical documentation
- B.6.2.8 AI system recording of event logs

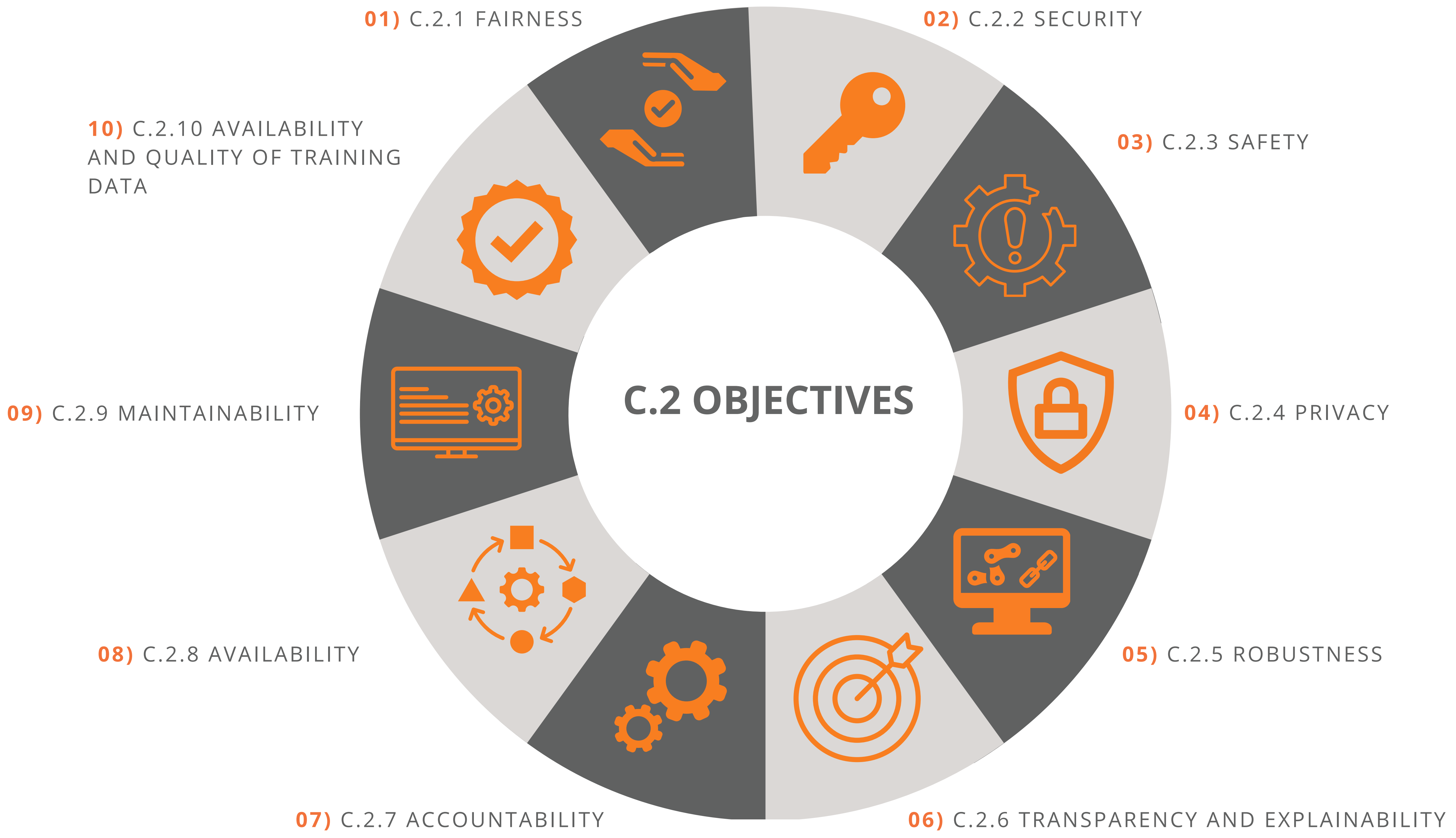
B.7 Data for AI systems
- B.7.2 Data for development and enhancement of AI system
- B.7.3 Acquisition of data
- B.7.4 Quality of data for AI systems
- B.7.5 Data provenance
- B.7.6 Data preparation

B.8 Information for interested parties of AI
- B.8.2 System documentation and information for users
- B.8.3 Understand-ability of provided information
- B.8.4 External reporting
- B.8.5 Communication of incidents
- B.8.6 Information for interested parties

B.9 Use of AI systems
- B.9.2 Processes for responsible
- B.9.3 Objectives for responsible use of AI system
- B.9.4 Intended use of the AI system

B.10 Third party relation
- B.10.2 Allocating responsibilities
- B.10.3 Suppliers
- B.10.4 Customers

ISO/IEC 42001 ANNEX C: Objectives and Risk Sources



C.2 OBJECTIVES

01) C.2.1 FAIRNESS

02) C.2.2 SECURITY

03) C.2.3 SAFETY

04) C.2.4 PRIVACY

05) C.2.5 ROBUSTNESS

06) C.2.6 TRANSPARENCY AND EXPLAINABILITY

07) C.2.7 ACCOUNTABILITY

08) C.2.8 AVAILABILITY

09) C.2.9 MAINTAINABILITY

10) C.2.10 AVAILABILITY AND QUALITY OF TRAINING DATA

C.3 RISK SOURCES

C.3.1 LEVEL OF AUTOMATION

C.3.2 LACK OF TRANSPARENCY AND EXPLAINABILITY

C.3.3 COMPLEXITY OF ENVIRONMENT

C.3.4 SYSTEM LIFE CYCLE ISSUES

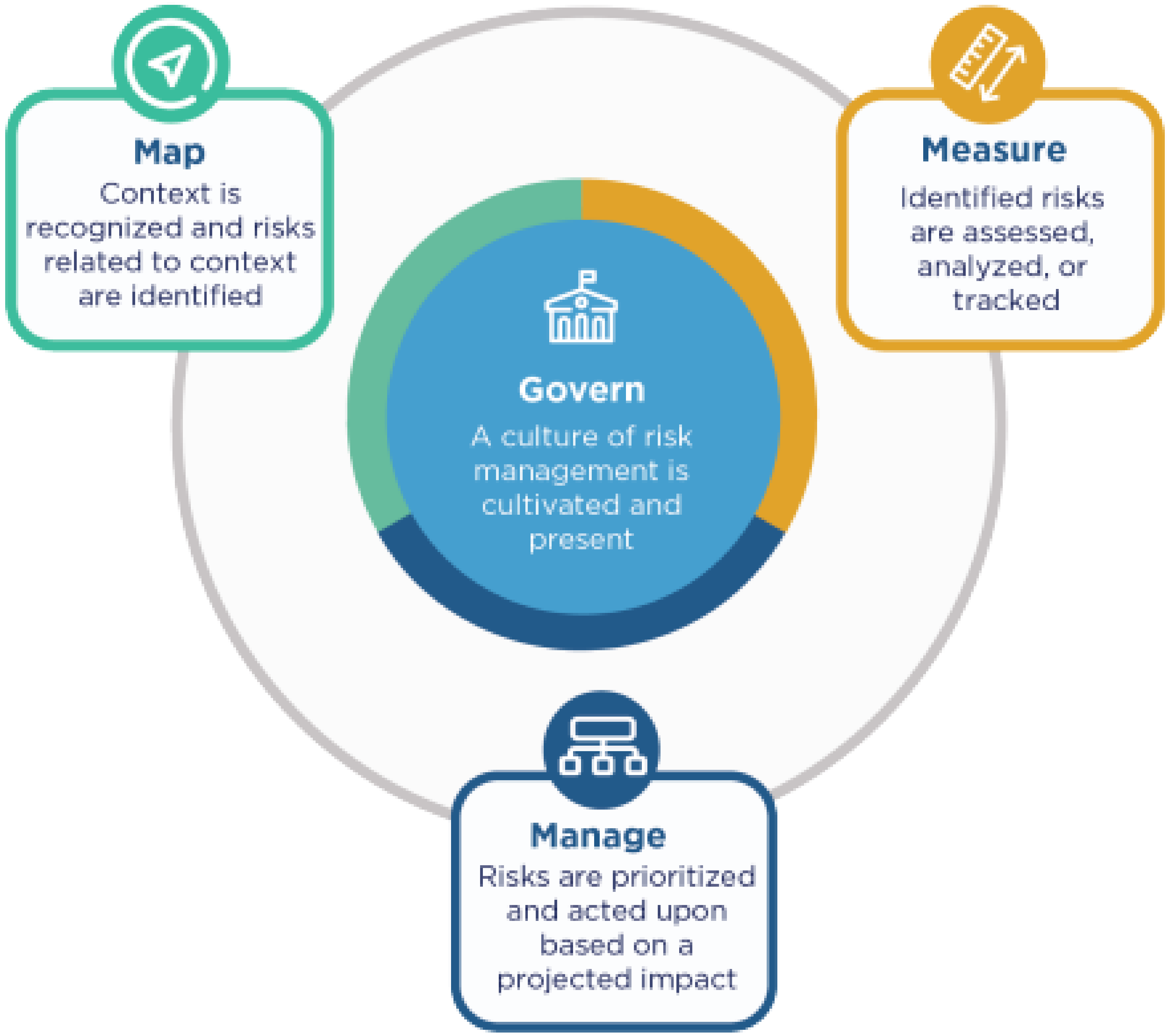
C.3.5 SYSTEM HARDWARE ISSUES

C.3.6 TECHNOLOGY READINESS

C.3.7 RISK SOURCES RELATED TO MACHINE LEARNING

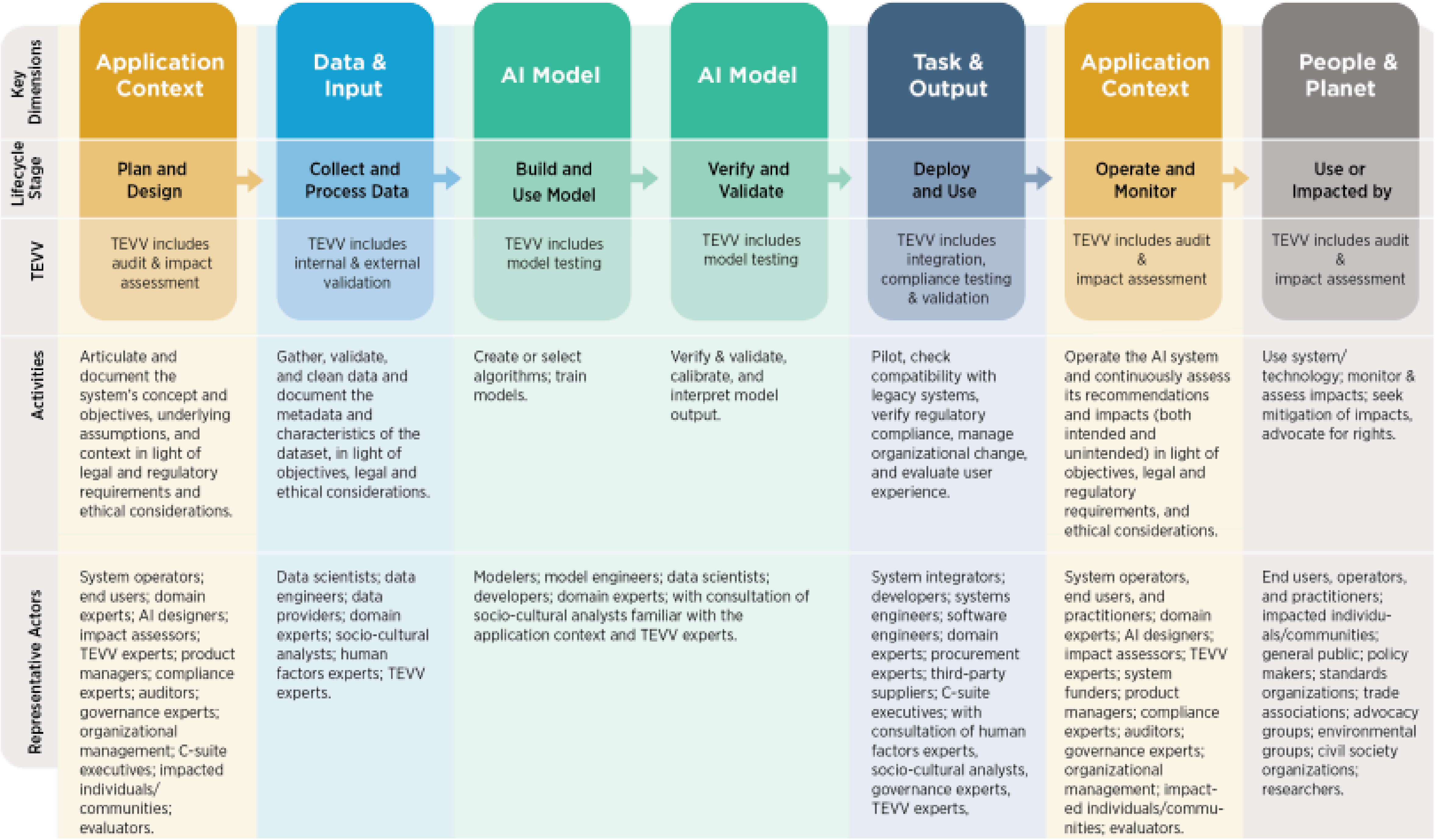
NIST AI RMF Core

The NIST AI RMF Core provides outcomes and actions that enable dialogue, understanding, and activities to manage AI risks and responsibly develop trustworthy AI systems. As illustrated in Figure the Core is composed of four functions: **GOVERN, MAP, MEASURE,** and **MANAGE**. Each of these high-level functions is broken down into categories and subcategories. Categories and subcategories are subdivided into specific actions and outcomes. Actions do not constitute a checklist, nor are they necessarily an ordered set of steps.


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

AI actors across AI lifecycle stages

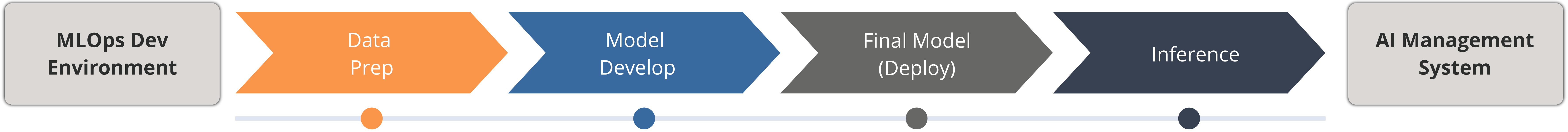
This overview delineates the roles and responsibilities associated with different stages of AI lifecycle management. Each stage, from planning and design through deployment and monitoring, involves specific tasks such as data collection, model building and verification . It is critical to differentiate between those who develop and use the models and those responsible for their verification and validation, ensuring unbiased and rigorously tested AI solutions.



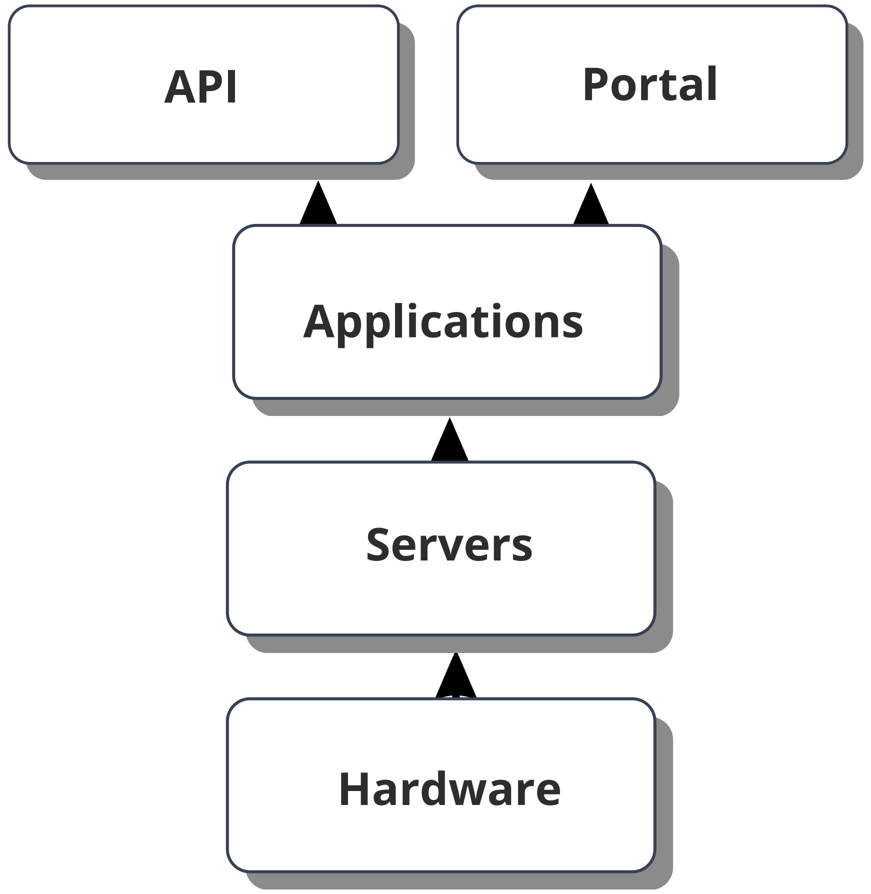
Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impacted individuals/communities; evaluators.	End users, operators, and practitioners; impacted individuals/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

OBTAINING ASSURANCE

DEVELOPMENT STACK



DEV ENVIRONMENT STACK



ASSURANCE REPORTING

COMPUTING PLATFORM

(SOC 2)

- Security
- Availability
- Data Confidientially
- Data Privacy

ALGORITHM ASSURANCE

(SOC 2 Processing Integrity)

- Model Specification and Processing
- Fair Model Treatment
- CI/CD w/ Continuous Training

GOVERNANCE

ISO/IEC 42001 AIMS

- Fairness
- Accountability
- Transparency and Explainability
- Security and Privacy
- Safety and Health
- Financial Consequences
- Accessibilty
- Human RIghts

Establishing Trustworthy AI

Assessment and Gap Analysis

Sigma Technology can assist organizations with thorough assessments and gap analysis to evaluate the organization's current AI practices against the requirements stipulated in ISO 42001. This involves:

- Identifying the scope of AI implementation within the organization.
- Assessing the existing AI systems, processes, and data usage against ISO 42001 guidelines.
- Documenting discrepancies and identifying areas for improvement to achieve compliance.

Development of AI Governance Framework

We assist organizations in developing a robust AI governance framework aligned with ISO 42001 principles. This includes

- Establishing policies and procedures for the responsible development, deployment, and monitoring of AI systems.
- Implementing mechanisms for accountability, transparency, and fairness in AI decision-making processes.
- Integrating risk management strategies to mitigate potential ethical, legal, and societal risks associated with AI technologies.

Continuous Monitoring and Improvement

We assist organizations in establishing mechanisms for continuous monitoring and improvement of their AI systems in line with ISO 42001. This involves

- Implementing ongoing monitoring and auditing processes to assess the performance and ethical implications of AI systems.
- Regularly updating AI governance frameworks and policies to reflect emerging ethical standards and regulatory requirements.
- Facilitating continuous learning and adaptation to ensure that AI technologies evolve responsibly alongside changing societal expectations and technological advancements

OFFERING

Sigma Technology offers comprehensive support to organizations seeking ISO 42001 compliance, ensuring that their AI initiatives adhere to the established standards and best practices. Our approach is structured to address key elements outlined in ISO 42001, facilitating a seamless integration of AI technologies while prioritizing ethical considerations and risk mitigation.

SIGMA TECHNOLOGY

ABOUT US

Sigma Technology Partners stands as an organization specializing in enterprise IT and cybersecurity solutions. Our range of services encompasses compliance consulting, cybersecurity, and Managed Security Provider (MSP) provisions, primarily catering to governmental and public sector entities. Our expertise extends to a comprehensive array of offerings, including assistance with AICPA SOC-2 Audit, FISMA compliance, FedRAMP Readiness Assessment, ISO/IEC 27001 Compliance consulting, Threat and Vulnerability Assessment, Cloud Architecture Assessment, and Penetration Testing Services.

(800)748-6602

info@sigmatechllc.com
https://www.sigmatechllc.com

Sigma Technology Partners LLC
2300 Wilson Blvd, #700 Arlington, VA 22201