Overview & major standardization achievements of the International Standardization Committee
ISO/IEC JTC 1/SC 42 Artificial Intelligence

Peter Deussen

peter.deussen@microsoft.com

SC 42 – Artificial Intelligence

Agenda

- Scope
- SC 42's role
- Participation
- Growth
- Outreach
- Collaboration with CEN-CENELEC/JTC 21

Overview

- Published standards
- Ongoing projects
- Work under preparation

Work programme

- Working groups
- Ad-hoc and Advisory Groups

Structure

- Concepts and terminology
- Overview of ethical and societal concerns
- Data quality for analytics and machine learning (ML)

Highlights

SC 42 Overview

The Role of International Standards in Artificial Intelligence (AI)

ISO/IEC JTC 1/SC 42
Artificial Intelligence

Scope

Standardization in the area of Artificial Intelligence

- *Serve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence*

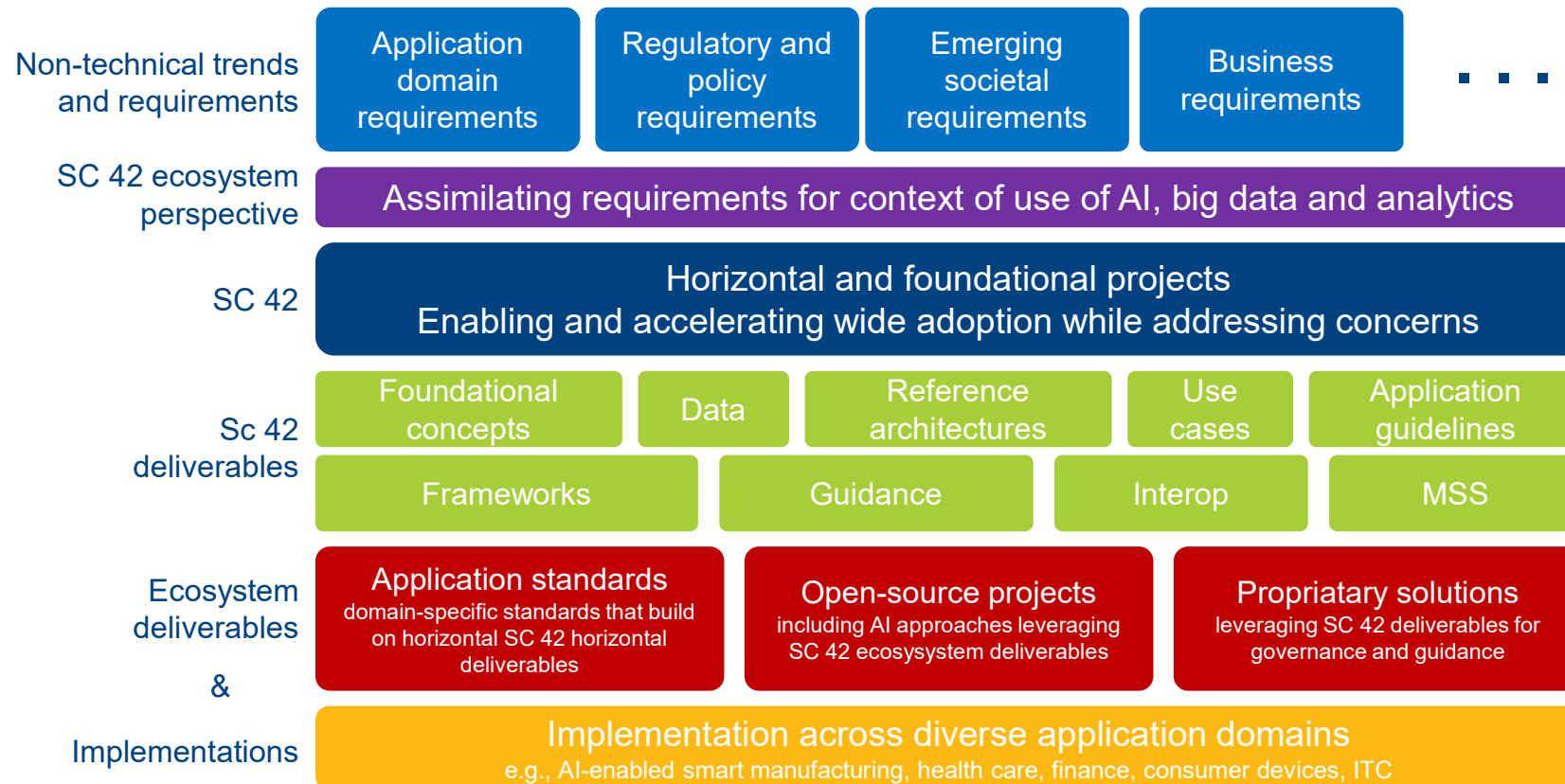- *Provide guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications*

Established: 2017, Beijing

Committee Manager: Heather Benko (ANSI)
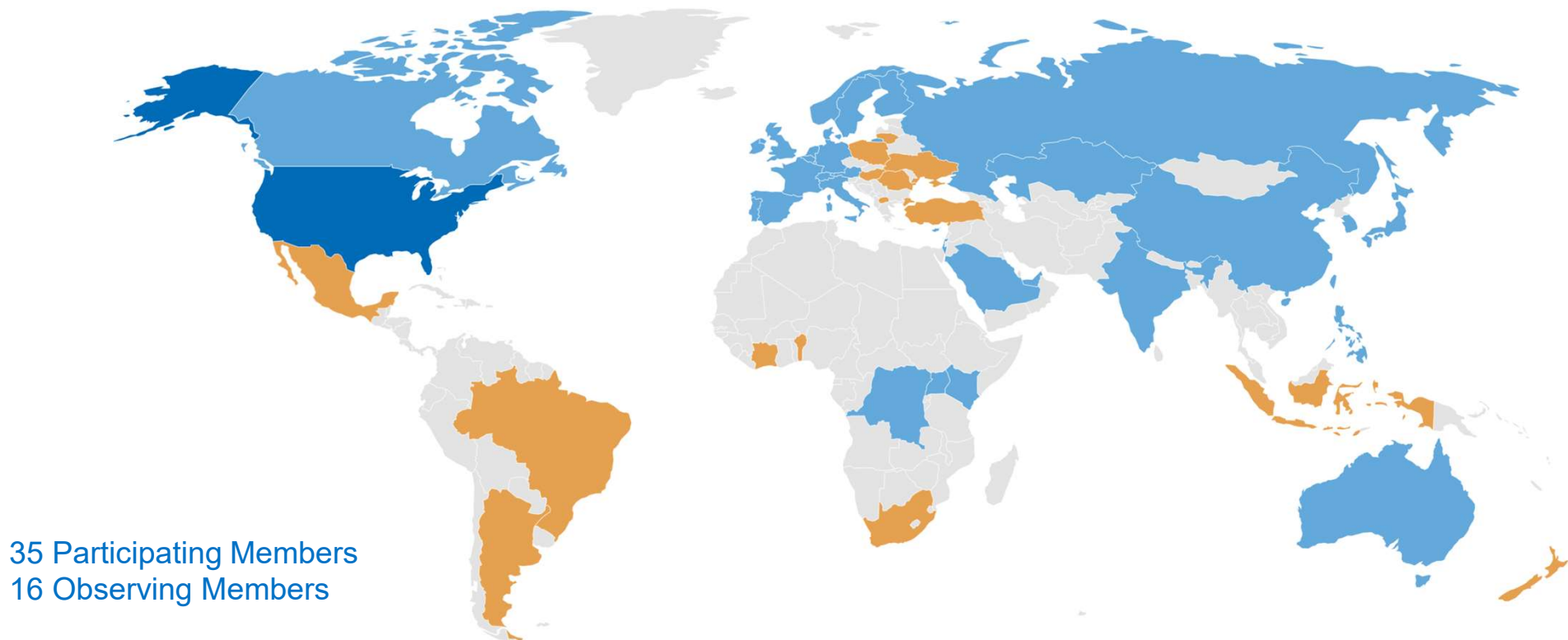
Chairperson: Wael William Diab (US)

SC 42's role
An Ecosystem Approach

Non-technical trends and requirements

| Application domain requirements | Regulatory and policy requirements | Emerging societal requirements | Business requirements | . . . |

SC 42 ecosystem perspective

Assimilating requirements for context of use of AI, big data and analytics

SC 42

Horizontal and foundational projects
Enabling and accelerating wide adoption while addressing concerns

Sc 42 deliverables

| Foundational concepts | Data | Reference architectures | Use cases | Application guidelines |

| Frameworks | Guidance | Interop | MSS |

Ecosystem deliverables

&

Implementations

| Application standards domain-specific standards that build on horizontal SC 42 horizontal deliverables | Open-source projects including AI approaches leveraging SC 42 ecosysystem deliverables | Propriatary solutions leveraging SC 42 deliverables for governance and guidance |

Implementation across diverse application domains
e.g., AI-enabled smart manufacturing, health care, finance, consumer devices, ITC

SC 42 Achievements
World-wide participation

35 Participating Members
16 Observing Members

Growth by the numbers

Projects: 15 published/ 25 active

Participation

- 51 NBs (35P/16O) despite economic headwinds

- 250 people at virtual

- 50+ liaisons (13 Cat A)

Growth in the diversity of the work program and ecosystem coverage: Extension of existing areas across including:

• MSS related projects to enable certification/audit • transparency taxonomy	• data quality governance; • treatment of unwanted bias	• oversight of AI systems • use cases revision	• ML computing devices • guidance for mitigating ethical issues

New areas include AI Testing, Verification and Validation (jointly with SC 7); AI-enabled Health Informatics (jointly with ISO TC 215)

Outreach

- Over a dozen IEC articles and multimedia engagements

- Collaboration with the ISO Comms on key articles and outreach plans

- Support of IEC and ISO AI related initiatives such as IEC Workshop on AI with Trust and ISO

- Progress on existing and new collaborations: IEC SC 65A, ISO 204, 215, SMCC, JTC 1 (SC 7, 27, 29, 32, 38, 39, 40, 41), OECD, EC, UN

- Launch of ISO/IEC Workshop Series

  - Bi-annual workshop series bringing AI luminaries to discuss AI ecosystem including emerging trends, technology, requirements, applications and the role of standards.

  - May inaugural workshop had 500 registrations per session and avg 300 attendance.

  - Registration for second workshop in November is open

ISO/IEC AI Workshop Series

Second Workshop on 29th and 30th November 2022 (Online)

4 Content Tracks

AI Applications – Healthcare

This track will look at emerging AI applications, public projects, use cases and case studies with the goal of identifying insights relating to AI application requirements, providing an overview of supporting horizontal standardization work and discuss roadmaps of application domain verticals.

This installment of the workshop will focus on AI in Healthcare applications

Beneficial AI

AI promises to change how we live, work and play for the better. Moreover, some applications of AI stand to have an unprecedent societal impact. This track will focus on advances to humanity that would otherwise not be possible without AI

Novel AI Standardization Approaches (Horizontal and Vertical)

In addition to building on well established information and operational technologies, standards and open-source efforts, AI introduces technology-specific challenges through its learning nature. This track will look at such challenges and discuss the innovative standards approaches to address them

Emerging AI Technology Trends and Requirements

AI technology and capability is evolving at a rapid pace. This track will look the combination of emerging technology trends and requirements through a diverse set of perspectives ranging from those luminaries driving innovation to those looking to ensure responsible adoption

Benefits

Innovation
workshop will focus on state of the art, priorities and requirements

Market Intelligence
identify emerging trends, requirements, insights and opportunities

Thought Leadership
participating in this interactive event will help to position your organization as an industry leader

Enhanced
Enhanced networking opportunities

Stakeholders
Attract new stakeholders to join standardization activities

Information
Gain insight into alternative POVs and needs through Q&A and panel discussion

Registration: https://jtc1info.org/technology/subcommittees/ai/workshops/w2/

Collaboration with CEN/CENELEC JTC 21 "Artificial Intelligence"

The top right has logos: ISO JTC1 IEC, INFORMATION TECHNOLOGY STANDARDS, SC 42 – Artificial Intelligence

SC 42 – Artificial Intelligence

Vienna Agreement will be applied for the collaboration between the two committees

JTC 21 is about to adopt various SC 42 deliverables, including

- ISO/IEC 22989:2022 – Concepts and terminology

- ISIEC 23053:2022 – Machine learning framework

- ISO/IEC 42001:- - Management system

- ISO/IEC 5259 Series on data quality

- . . .

SC 42 is about to establish an AHG to understand which JTC 21 projects are candidates for parallel development according to the VA

Under discussion: How can SC 42 work be leveraged to address the Standardization requests of the European Commission in the context of the Draft AI Act

SC 42 Work programme overview

- **Published standards**

- **Standards under development**

- **Work under preparation**

Published standards

Foundational

- ISO/IEC 22989 AI concepts and terminology

- ISO/IEC 23053 Framework for AI Systems using ML

Trustworthiness

- ISO/IEC TR 24028 Overview of trustworthiness in artificial intelligence

- ISO/IEC TR 24027 Bias

- ISO/IEC 38507 Governance (jointly developed with SC 40)

- ISO/IEC TR 24368 Ethics

- ISO/IEC TR 24029-1 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview

Use cases and applications

- ISO/IEC TR 24030 Use cases

Data

- ISO/IEC 20546 Big data — Overview and vocabulary

- ISO/IEC TR 20547 Big data reference architecture
  - Part 1: Framework and application process
  - Part 2: Use cases and derived requirements
  - Part 3: Reference architecture
  - Part 5: Standards roadmap

Computational

- ISO/IEC 24372 Overview

- ISO/IEC TS 4213:2022 Assessment of machine learning classification performance

Standards under development

- ISO/IEC CD 5259-1 Data quality for analytics and machine learning (ML)
  - Part 1: Overview, terminology, and examples
  - Part 2: Data quality measures
  - Part 3: Data quality management requirements and guidelines
  - Part 4: Data quality process framework
  - Part 5: Data quality governance
- ISO/IEC DIS 5338 AI system life cycle processes
- ISO/IEC CD 5339 Guidelines for AI applications
- ISO/IEC CD 5392 Reference architecture of knowledge engineering
- ISO/IEC CD TR 5469 Functional safety and AI systems
- ISO/IEC AWI TS 5471 Software and systems engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of AI systems
- ISO/IEC AWI TS 6254 Objectives and approaches for explainability of ML models and AI systems
- ISO/IEC DIS 8183 Data life cycle framework
- ISO/IEC DIS 25059 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

- ISO/IEC AWI TS 8200 Controllability of automated artificial intelligence systems
- ISO/IEC AWI TS 12791 Treatment of unwanted bias in classification and regression machine learning task
- ISO/IEC AWI 12792 Transparency taxonomy of AI systems
- ISO/IEC AWI TS 17847 Verification and validation analysis of AI systems
- ISO/IEC AWI TR 17903 Overview of machine learning computing devices
- ISO/IEC FDIS 23894 Guidance on risk management
- ISO/IEC DIS 24029-2 Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods
- ISO/IEC AWI TR 24030 Use cases
- ISO/IEC 24668 Process management framework for big data analytics
- ISO/IEC AWI TS 29119-11 Testing for AI systems
- ISO/IEC DIS 42001 Management system
- ISO/IEC AWI 42005 AI system impact assessment

SC 42 Achievements
Work under preparation

- Currently under combined NP ballot/CD consultation:

  - ISO/IEC 42006 Requirements for bodies providing audit and certification of artificial intelligence management systems

    - Editor: Susanne Kuch (DE)

    - Explains how to do conformity assessment for ISO/IEC 42001, competencies needed for auditors, time needed for audit

- New technical report

  - Environmental sustainability aspects of AI

SC 42 Structure

- **Working groups**

- **Ad-hoc and advisory groups**

SC 42 Groups

Working Groups

WG 1	Foundational standards	Paul Cotton (CA)
WG 2	Data	Wo Chang (IE)
WG 3	Trustworthiness	David Filip (IE)
WG 4	Use cases and applications	
WG 5	Computational approaches and computational characteristics of AI systems	Ning Sun (CN)
JWG 2	Testing of AI systems with SC 7	Adam Leon Smith (SC 42/UK); Stuart Reid (SC 7/UK)
JWG 3	AI-enabled Health Informatics with ISO TC 215	TBD

Ad-hoc and Advisory Groups

AG 3	AI standardization roadmapping	Patrick Bezombes (FR)
AHG 4	Liaison with SC 27	Peter Deussen (DE)
AHG 6	Considering logistics relative to Comment Resolutions – CD/DIS Ballots	Paul Cotton (CA)
AHG 7	Considering potential JTC 21 projects for joint development under the Vienna Agreement	TBD

SC 42 Highlights

- **ISO/IEC 22989 Concepts and terminology**

- **ISO/IEC TR 24368:2022 Overview of ethical and societal concerns**

- **ISO/IEC 5259 Series on Data quality for analytics and machine learning (ML)**

ISO/IEC 22989 Concepts and terminology

SC 42 – Artificial Intelligence

Scope

This document establishes terminology for AI and describes concepts in the field of AI.

This document can be used in the development of other standards and in support of communications among diverse, interested parties or stakeholders.

This document is applicable to all types of organizations (e.g. commercial enterprises, government agencies, not-for-profit organizations).

Editor: Wei Wei (DE)

Current state:

- Published

- Defect Editing Group established to address editorial problems

ISO/IEC 22989 Concepts and terminology

Main contents

- Definitions of concepts relevant to AI
- General Concepts
- Machine learning concepts
- Examples of machine learning algorithms
- Autonomy, heteronomy and automation
- Internet of things and cyber-physical systems
- Trustworthiness
- AI verification and validation
- Jurisdictional issues
- Societal impact

- AI stakeholder roles
- AI system life cycle
- AI system functional overview
- AI ecosystem
- Fields of AI
- Applications of AI systems
- Mapping of the AI system lifecycle with the OECD's definition of AI system

ISO/IEC 22989 Concepts and terminology

SC 42 – Artificial Intelligence

AI definition

artificial intelligence

AI

*<discipline> research and development of mechanisms and applications of **AI systems***

artificial intelligence system

AI system

engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

- Result of a 4 year's discussion, highly controversial and disputed up to the last minute

- No consensus could be reached on a definition of AI as system property/capability

- Alignment with OECD, EU definitions was intended (as they were available during development of the project)

SC 42 Highlights
ISO/IEC 22989 Concepts and terminology

AI system lifecycle

Intended to be high-level to provide a template that can be used for other standardization work

Inception
- Objectives
- Requirements
- Risk assessment
- Risk treatment plan
- Polices and compliance

Design and development
- Approach
- Architecture
- Code
- Training and validation data
- Risk treatment

Verification and validation
- Test data
- System verification
- Risk monitoring and review

Re-evaluate
- Evaluate operating results
- Refine objectives
- Refine requirements
- Risk monitoring and review

Operation and monitoring
- Operating data input
- Model execution
- Model update
- Risk management

Deployment
- Runtime deployment
- Model deployment
- Risk treatment

Retirement
- Data disposal
- Model disposal
- Decommission and discard

Continuous validation
- Test data
- System validation
- Risk management
- Continuous improvement

If using continuous learning

ISO/IEC TR 24368:2022 Overview of ethical and societal concerns

Scope

This document provides a high-level overview of AI ethical and societal concerns.

In addition, this document;

- *provides information in relation to principles, processes and methods in this area;*

- *is intended for technologists, regulators, interest groups, and society at large;*

- *is not intended to advocate for any specific set of values (value systems).*

This document includes an overview of International Standards that address issues arising from AI ethical and societal concerns.

Editor: Viveka Bonde (SE)

Current state: Published

SC 42 – Artificial Intelligence

ISO/IEC TR 24368:2022 Overview of ethical and societal concerns

- Fundamental sources
- Ethical frameworks
- Human rights practices
- Themes and principles
- Examples of practices for building and using ethical and socially acceptable AI
  - Aligning internal process to AI principles
  - Considerations for ethical review framework
- Considerations for building and using ethical and socially acceptable AI
- Non-exhaustive list of ethical and societal considerations

- AI principles documents
  - Berkman Klein report
  - Global
  - North America
  - South America
  - Africa
  - Europe
  - Asia
  - Eurasia
  - Asia Pacific
- Use case studies

SC 42 – Artificial Intelligence

ISO/IEC 5259 Series on Data quality for analytics and machine learning (ML)

Number	Title	Editor	Current state
5259-1	Overview, terminology, and examples	Suwook Ha (KR)	CD
5259-2	Data quality measures	Kyoung-Sook Kim (JP)	CD
5259-3	Data quality management requirements and guidelines	Martin Saerbeck (DE)	CD
5259-4	Data quality process framework	Whanzong Ma (CN)	CD
5259-5	Data quality governance	Gyeung-Min Kim (KR)	WD

ISO/IEC 5259 Series on Data quality for analytics and machine learning (ML)

SC 42 – Artificial Intelligence

Data management life cycle and data quality management
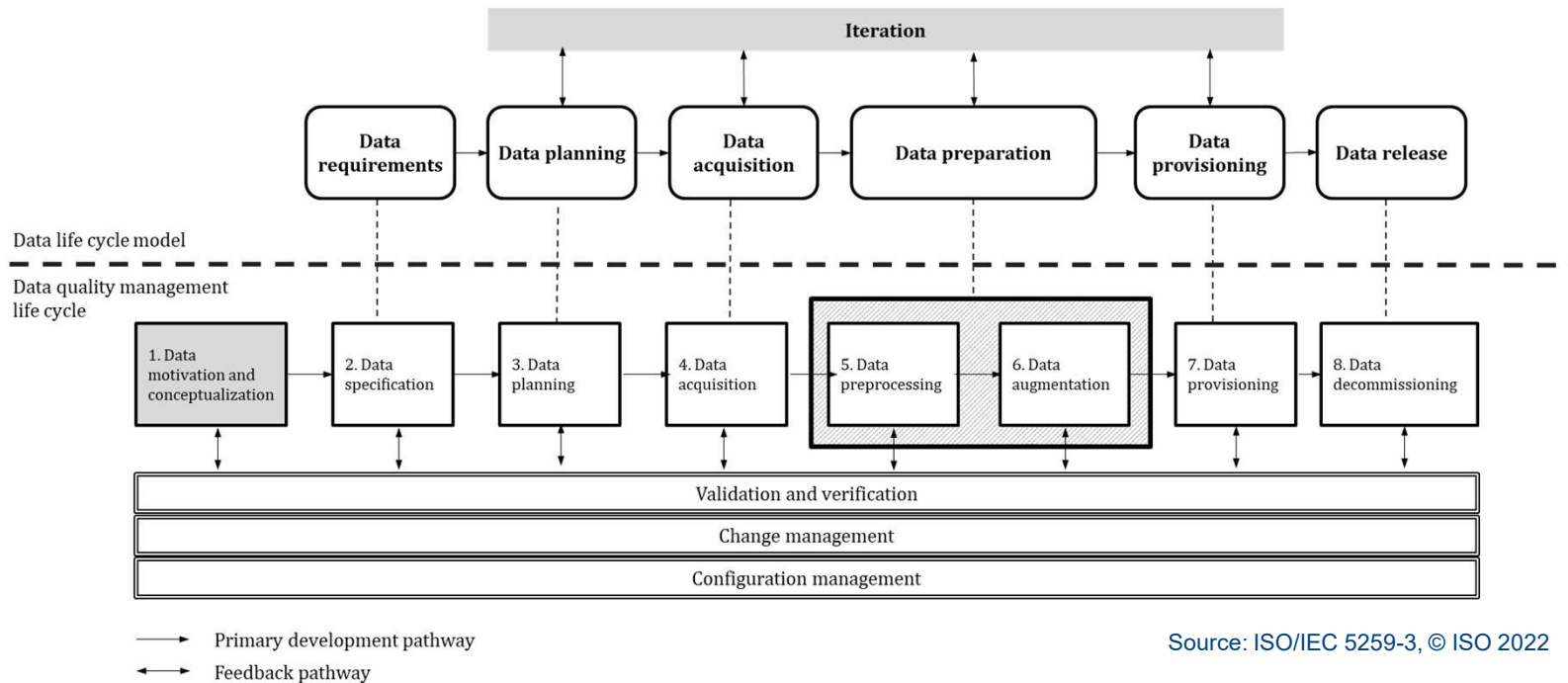
Source: ISO/IEC 5259-3, © ISO 2022

The Role of International Standards in Artificial Intelligence (AI)

THANKS !   QUESTIONS ?   REMARKS ?

Key Topics for Organizations/users using AI Technology

ISO/IEC 42001 – AI management system

Peter Deussen

peter.deussen@microsoft.com

SC 42 – Artificial Intelligence

Agenda

ISO/IEC 42001 Artificial Intelligence Management System (AIMS)

AIMS audit and certification

Scope

Structure

Example on for requirements, controls and implementation guidance

ISO/IEC 42001 Artificial Intelligence Management System (AIMS)

ISO management system standards (MSS) help organizations

- to improve their performance by specifying repeatable steps that organizations consciously implement to achieve their goals and objectives

- to create an organizational culture that reflexively engages in a continuous cycle of self-evaluation, correction and improvement of operations and processes through heightened employee awareness and management leadership and commitment

AIMS is a

- Management system standard on the **responsible development and use of AI**

- Comparable with ISO 9001, ISO/IEC 27001, etc.

- Highly strategic: High political impact and market relevance expected

ISO/IEC 42001
AIMS will be an auditable and certifiable standard!

- Audits are a vital part of the management system approach as they enable a company or organization to check how well their achievements meet their objectives and show conformity to the standard.

- ISO recognises different forms of assessment:
  - 1st party – the same organisation
  - 2nd party – customer/supplier in a relationship with the organisation
  - 3rd party – independent organisation, trusted by both parties

- ISO calls this conformity assessment: https://www.iso.org/conformity-assessment.html

ISO/IEC 42001 – Artificial Intelligence Management System (AIMS)

Scope

*This document specifies the **requirements and provides guidance for establishing, implementing, maintaining and continually improving an AI management system** within the context of an organization.*

*This document is intended for use by an organization providing or using products or services that utilize AI systems. This document helps the organization **develop or use AI systems responsibly in pursuing its objectives and meet applicable regulatory requirements, obligations related to interested parties and expectations from them**.*

This document is applicable to any organization, regardless of size, type and nature, that provides or uses products or services that utilize AI systems.

Editor: Marta Janczarski (IE)

Current status:

- DIS

- Publication expected mid 2023

ISO/IEC 42001
Overall Structure

Clause 1: Scope

Clause 2: Normative references

Clause 3: Terminology

Clause 4: Abbreviations

Management part (Clause 5 – 10)

Requirements on how to manage AI system in a responsible way

- Mandatory for compliance

- Structure and template requirements provided by ISO/IEC JTC 1 Supplement (Common Text)

- AIMS adopts the Common Text with minimal changes to maintain compatibility with other MSS

Control part (Annex A) and implementation guidance (Annex B)

Technical and organizational measures in support of management requirements

- Optional for compliance, but organizations have to provide rationale if a control is not implemented

- Described in terms of

  - Objectives

  - Controls to achieve an objective

  - Implementation guidance

ISO/IEC 42001
Management part - Requirements

Context of the organisation	Leadership	Planning	Support	Operation	Performance evaluation	Improvement
Understanding the organisation and its context	Leadership and commitment	Actions to address risks and opportunities	Resources	Operational planning and control	Monitoring, measurement, analysis and evaluation	Non-conformity and corrective action
Understanding the needs and expectations of interested parties	Policy	AI system impact assessment	Competence	AI risk assessment	Internal audit	Continual improvement
Determining the scope of the management system	Organisational roles, responsibilities and authorities	Management system objectives and planning to achieve them	Awareness	AI risk treatment	Management review	
The management system			Communication	AI system impact assessment		
			Documented information			

ISO/IEC 42001
Controls and implementation guidance

Policies related to AI	Internal organization	Resources for AI systems	Assessing impacts of AI systems	AI system life cycle	AI system development life cycle	Data for AI systems	Information for interested parties	Use of AI systems	Third party relationships
AI policy	AI roles and responsibilities	Resource documentation	AI system impact assessment process	Management guidance for AI system development	AI system requirements and specification	Data for development and enhancement of AI system	System documentation and information for users	Processes for responsible use of AI	Allocating responsibilities
Alignment with other organizational policies	Reporting of concerns	Data resources	Documentation of AI system impact assessments	Objectives for responsible development of AI system	Documentation of AI system design and development	Acquisition of data	Understandability and accessibility of provided information	Objectives for responsible use of AI system	Suppliers
Review of the AI policy		Tooling resources	Assessing AI system impact on individuals and groups of individuals	Processes for responsible design and development of AI systems	AI system verification and validation	Quality of data for AI systems	External reporting	Intended use of the AI system	Customers
Review of the AI policy		System and computing resources	Assessing societal impacts of AI systems		AI system deployment	Data provenance	Communication of incidents		
		Human resources			AI system operation and monitoring	Data preparation	Information for interested parties		
					AI system technical documentation				
					AI system recording of event logs				

ISO/IEC 42001
Requirements and controls

Example: AI system impact assessment

Requirement

The organization **shall** assess the potential consequences for individuals and societies that can result from the development or use of AI systems. The AI system impact assessment **shall** determine the potential consequences an AI system's deployment and intended use has on individuals and societies. The result of the system impact assessment **shall** be documented and be made available to relevant interested parties where appropriate.

The organization **should** consider whether an AI system affects:

– the legal position or life opportunities of individuals;

– the physical or psychological well-being of individuals;

– universal human rights;

– society.

The organization **can** integrate the AI system impact assessment into their risk assessment (Clause 6.1.2).

Objective

To assess AI system impacts to individuals and societies affected by the AI system throughout its life cycle.

Control

The organization should assess the potential consequences for individuals and societies that would result from the development or use of AI systems.

[. . .]

Implementation guidance

Topics the organization should consider can include, but are not limited to:

a)   circumstances under which an AI system impact assessment should be performed, can include

1.   criticality of the intended purpose and context in which the AI system is used or any significant changes to these;

2.   complexity of AI technology and the level of automation of AI systems or any significant changes to that;

3.   sensitivity of data types and sources processed by the AI system or any significant changes to that.

[…]

THANKS !   QUESTIONS ?   REMARKS ?

Key Topics for Organizations/users using AI Technology

ISO/IEC 23894 – Guidance on risk management

Peter Deussen

peter.deussen@microsoft.com

SC 42 – Artificial Intelligence

Agenda

AI risks

Risk management: ISO 31000 overview

ISO/IEC 23894 Guidance on risk management

Principles, framework and processes

Artificial Intelligence Risks

AI RISKS ARE RELATED TO

Fairness

Security

Safety

Privacy

Robustness

Transparency / explainability

Environmental impact

Accountability

Maintainability

Availability and quality of data

AI expertise

CAUSED BY

Level of automation

Lack of transparency and explainability

Complexity of environment

System life cycle issues

System hardware issues

Technology readiness

Machine learning issues
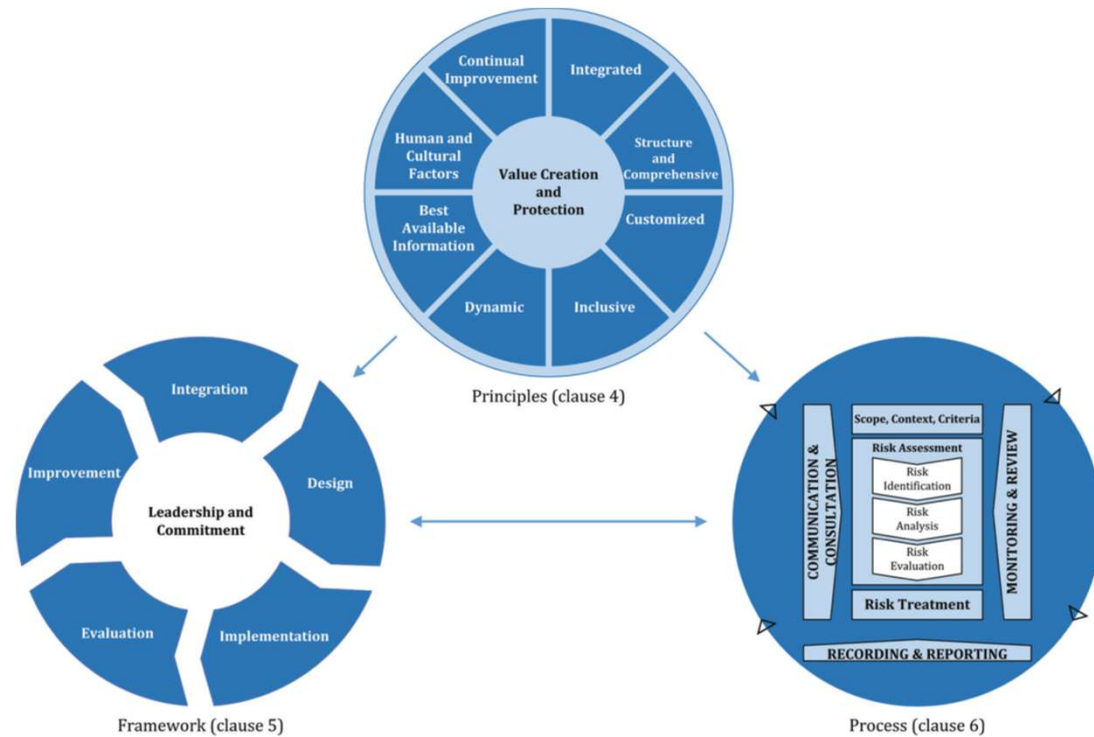
Risk Management- Overview

Risk management is described in ISO 31000:2018

- Well accepted in industry

Approach comprises of

- Principles to follow

- Management framework

- Processes (focus for today)

ISO/IEC 23894 follows this approach and adopts ISO 31000 entirely, but adds AI specific considerations

[Source: ISO/IEC 31000:2018, © ISO]

ISO/IEC 23894
Artificial Intelligence Risk Management

Scope

This document provides guidance on how organizations that develop, produce, deploy or use products, systems and services that utilize artificial intelligence (AI) can manage **risk specifically related to AI**. *The guidance also aims to* **assist organizations to integrate risk management into their AI-related activities and functions.** *It moreover describes* **processes for the effective implementation and integration** *of AI risk management.*

The application of these guidance can be customized to any organization and its context.

Editor: Peter Deussen (DE)

Current state: FDIS ballot about to start, publication expected for early 2023

- Please note:

  - Guidance only!
    - No requirements
    - No basis for certification

- Relates to risks to the organization

  - Impact to individuals, society, environment considered as part of organizational risk management

Principles of risk management

ISO 31000
PRINCIPLES

Integrated

Structured and comprehensive

Customized

Inclusive

Dynamic

Best available information

Human and cultural factors

Continual improvement

ISO/IEC 23894
AI SPECIFIC INTERPRETATION

Extended need to understand stakeholder needs and expectations

Nature of AI systems

Implications to the organization itself

Consideration of societal concerns

Risk management framework

ISO 31000 FRAMEWORK COMPONENTS

Leadership and commitment

Integration

Design

Implementation of risk management

Evaluation of risk management

Improvement of risk management

ISO/IEC 23894 AI SPECIFIC INTERPRETATION

Communicate responsible use / development of AI systems

Understand the context of the organization
- Social, cultural, political, legal, regulatory, financial, technological, economic and environmental factors
- key drivers and trends
- stakeholder relationships, perceptions, values, needs and expectations
- contractual relationships and commitments;
- complexity of networks and dependencies
- Internal factors, e.g., culture and values, organizational strucutres, capabilities, data

Define roles and responsibilities, allocate resource

Determine needed information for continual improvement

1. Define scope, Context, Risk Criteria

Scope	Context	Risk Criteria
• Objectives and expected outcomes • time, location, specific inclusions and exclusions; • tools and techniques; • resources and responsibilities • relationships with other projects, processes and activities.	• Societal, legal and contractual environment of the organization • Stakeholder identificaiton, their needs and expectectation	• the nature and type of uncertainties that can affect outcomes and objectives (both tangible and intangible) • how consequences and likelihood will be defined and measured • time-related factors • consistency in the use of measurements • how the level of risk is to be determined • how combinations and sequences of multiple risks will be taken into account • the organization's capacity

2. Risk Assessment

Risk identification	Risk analysis	Risk evaluation

• Identification of assets and their value:
  • Assets of and their value to the organization
  • Assets of and their value to individuals
  • Assets of and their value to society
• Identification of risk sources
• Identification of potential outcomes
• Identification of controls
• Identification of consequences
	• Assessment of consequences:
	  • criticality of the impact
	  • tangible and intangible impacts
	  • criteria used to establish the overall impact
	• Assessment of likelihood
		• Compare the results of the risk analysis with the established risk criteria to determine where additional action is required

3. Risk Treatment

Treatment options
• avoiding the risk by deciding not to start or continue with the activity
• taking or increasing the risk in order to pursue an opportunity
• removing the risk source
• changing the likelihood
• changing the consequences
• sharing the risk (for instance, through contracts or buying insurance)
• retaining the risk by informed decision

Additional steps

Communication and consultation	Monitoring and review	Recording and reporting
• Involvement and information of stakeholders • Communication of policies and commitments	• Determining the effectiveness of the risk management processes • Improvements	• Information to be recorded • Reporting requirements

THANKS !   QUESTIONS ?   REMARKS ?

Key Topics for Organizations/users using AI Technology

ISO/IEC 42005 – AI system impact assessment

Peter Deussen

peter.deussen@microsoft.com

SC 42 – Artificial Intelligence

Agenda

Motivation: Gap between ISO/IEC 42001 and ISO/IEC 23894
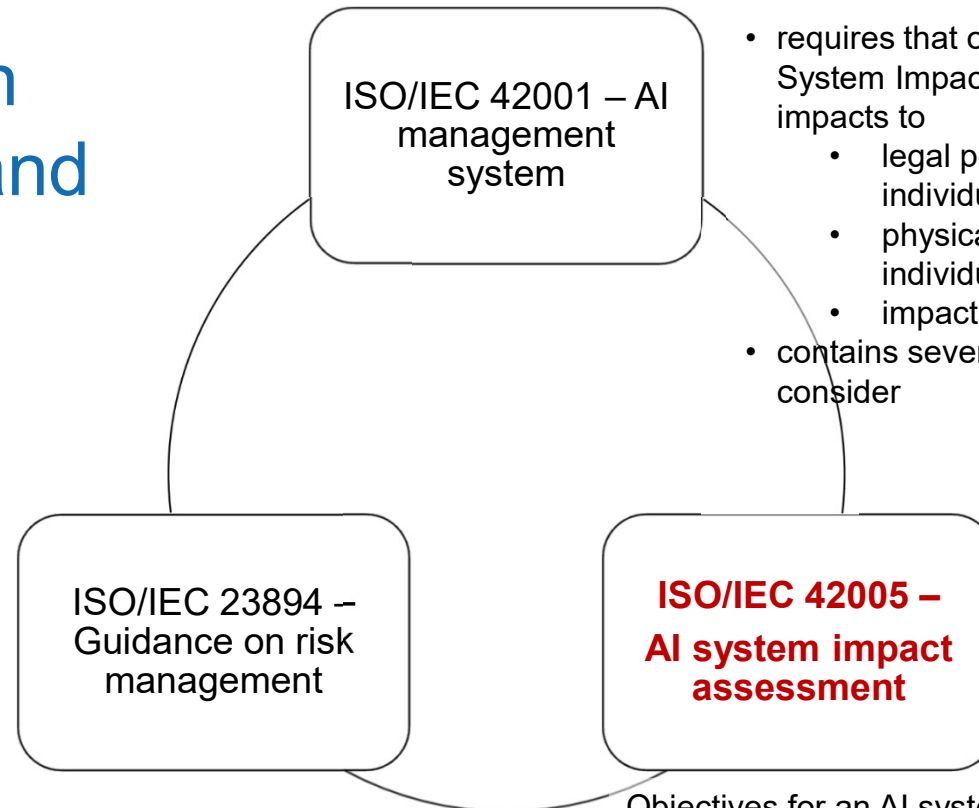
ISO/IEC 42005 AI system impact assessment

Main contents

The gap between ISO/IEC 42001 and ISO/IEC 23894

ISO/IEC 42001 – AI management system

- requires that organizations perform an AI System Impact Assessment, considering impacts to
  - legal position or life opportunities of individuals;
  - physical or psychological well-being of individuals;
  - impacts to human rights
- contains several guidelines on WHAT to consider

ISO/IEC 23894 – Guidance on risk management

- contains recommendations to take into account factors such as privacy, security, safety, fairness, transparency while managing organizational risks
- those factors are related to external stakeholders (individuals, society)

Further related to: ISO/IEC FDIS 38507 - Governance implications of the use of artificial intelligence by organizations

ISO/IEC 42005 – AI system impact assessment

Neither ISO/IEC 23894 nor ISO/IEC 42001 provides any guidelines on how to do perform the actual AI system impact assessment

Objectives for an AI system impact assessment guidance document include:
- providing organizations with guidance for achieving a key requirement of AIMS and assist in achieving responsible development and use of AI systems
- providing organizations with guidance that is necessary to institute appropriate AI risk management processes

ISO/IEC 42005
AI system impact assessment

Scope

*This document provides guidance for organizations performing **AI system impact assessments** for **individuals and societies** that can be affected by an AI system and its **foreseeable applications**. It includes considerations for how and when to perform such assessments and at what stages of the AI system lifecycle, as well as guidance for AI system impact assessment documentation.*

*Additionally, this guidance includes **how this AI system impact assessment process can be integrated into an organization's AI risk management and AI management system**.*

This document is intended for use by organizations developing, providing, or using AI systems. This document is applicable to any organization, regardless of size, type and nature.

Editor: Peter Deussen (DE)

Current state: WD

- Please note:
  - Guidance only!
    - No requirements
    - No basis for certification

ISO/IEC 42005
Main contents

Implementing an AI system impact assessment process

- Documenting the process

- Timing of AI system impact assessment

- Guidance for determining the scope of the impact assessment, domain of intended use

- Allocating responsibilities

- Integration into the AI system life cycle

- Establishing thresholds (e.g. sensitive uses, prohibited uses)

- Performing the impact assessment

- Analysing the results of the impact assessment

- Recording and reporting

- Approval process

- Monitoring and review

Documenting the AI system impact assessment

- Scope of the impact assessment

- System information

- System description

- System features

- System purposes and intended uses

- Algorithm information

- Data information and quality

- Deployment environment

- Geographical area, languages

- Stakeholders

- Potential impacts: Benefits and harms

- System failures and misuse

Measures to address harms and benefits

- No content available yet

- Not intended to provide full guidance on treatment measures (risk management), but concentrate on technical measures from a development and operational perspective

- Annexes

- Guidance for use with ISO/IEC 42001

- Guidance for use with ISO/IEC 23894

THANKS !   QUESTIONS ?   REMARKS ?

SC 42 – Artificial Intelligence

Peter Deussen

National Standards Officer Germany

Microsoft Deutschland GmbH

Mobile: +49 151 4406 3650

E-Mail: peter.deussen@microsoft.com