INTERNATIONAL STANDARD

ISO/IEC 5338

First edition
2023-12

Information technology — Artificial intelligence — AI system life cycle processes

Technologies de l'information — Intelligence artificielle — Processus de cycle de vie des systèmes d'IA

Reference number
ISO/IEC 5338:2023(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Artificial intelligence (AI) systems in the fields of computer vision and image recognition, natural language processing, fraud detection, automated vehicles, predictive maintenance and planning have achieved remarkable successes. To build and maintain an AI system, it is an efficient approach to extend the life cycle processes for a traditional software system to include AI-specific life cycle characteristics.

An example of such a specific characteristic of an AI system life cycle is where a system employs machine learning (ML) using training data and it becomes necessary to retrain the ML model using new training data that is more representative of current production data.

ISO/IEC/IEEE 12207 describes software life cycle processes and ISO/IEC/IEEE 15288 describes system life cycle processes. While these life cycle processes are broadly applicable to AI systems, they require the introduction of new processes and the modification of existing processes to accommodate the characteristics of AI systems. This document extends the current generic life cycle process International Standards to make them applicable for AI systems so that the AI system life cycle can benefit from established models and existing practices. Some AI systems are in use in areas which are related to safety, such as health care or traffic control. Such safety critical AI systems need special attention and considerations as described in ISO/IEC TR 5469 [5].

Integrating the AI system life cycle into existing processes delivers efficiency gains, better adoption of AI and mutual understanding among AI system stakeholders as defined in ISO/IEC 22989. Such an integrated life cycle approach embraces the fact that AI systems typically are a combination of AI-specific elements and traditional elements such as source code and databases.

This document provides further details on AI system life cycle processes as discussed in ISO/IEC 42001[18].

Information technology — Artificial intelligence — AI system life cycle processes

1 Scope

This document defines a set of processes and associated concepts for describing the life cycle of AI systems based on machine learning and heuristic systems. It is based on ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207 with modifications and additions of AI-specific processes from ISO/IEC 22989 and ISO/IEC 23053.

This document provides processes that support the definition, control, management, execution and improvement of the AI system in its life cycle stages. These processes can also be used within an organization or a project when developing or acquiring AI systems. When an element of an AI system is traditional software or a traditional system, the software life cycle processes in ISO/IEC/IEEE 12207 and the system life cycle processes in ISO/IEC/IEEE 15288 can be used to implement that element.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC/IEEE 15288:2023, *Systems and software engineering — System life cycle processes*

ISO/IEC/IEEE 12207:2017, *Systems and software engineering — Software life cycle processes*

ISO/IEC 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC 23053, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989, ISO/IEC 23053, ISO/IEC/IEEE 15288, ISO/IEC/IEEE 12207 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
knowledge acquisition
process of locating, collecting, and refining knowledge and converting it into a form that can be further processed by a knowledge-based system

Note 1 to entry: Knowledge acquisition normally implies the intervention of a knowledge engineer, but it is also an important component of machine learning.

[SOURCE: ISO/IEC 2382:2015, 2123777, modified — Notes 2 to entry 3 to entry have been deleted.]

4 Abbreviated terms

AI artificial intelligence

ML machine learning

5 Key concepts

5.1 General

AI system life cycle consists of three types of processes:

— Generic processes: Processes that are identical to the processes defined in ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207.

— Modified processes: Processes where elements are modified, added or removed from the ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207 definition.

NOTE 1 The Clause for each of these “Modified processes” contains a subclause of AI-specific particularities that provide guidance to adapt the process to AI systems.

— AI-specific processes: Processes that are specific to characteristics of AI systems but are not based directly on any processes in ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207.

AI system life cycle processes in Clause 6 are presented as generic, modified or AI-specific. Figure 1 shows the life cycle processes of AI system, grouped by type, and compared to ISO/IEC/IEEE 15288:2023, Figure 4.

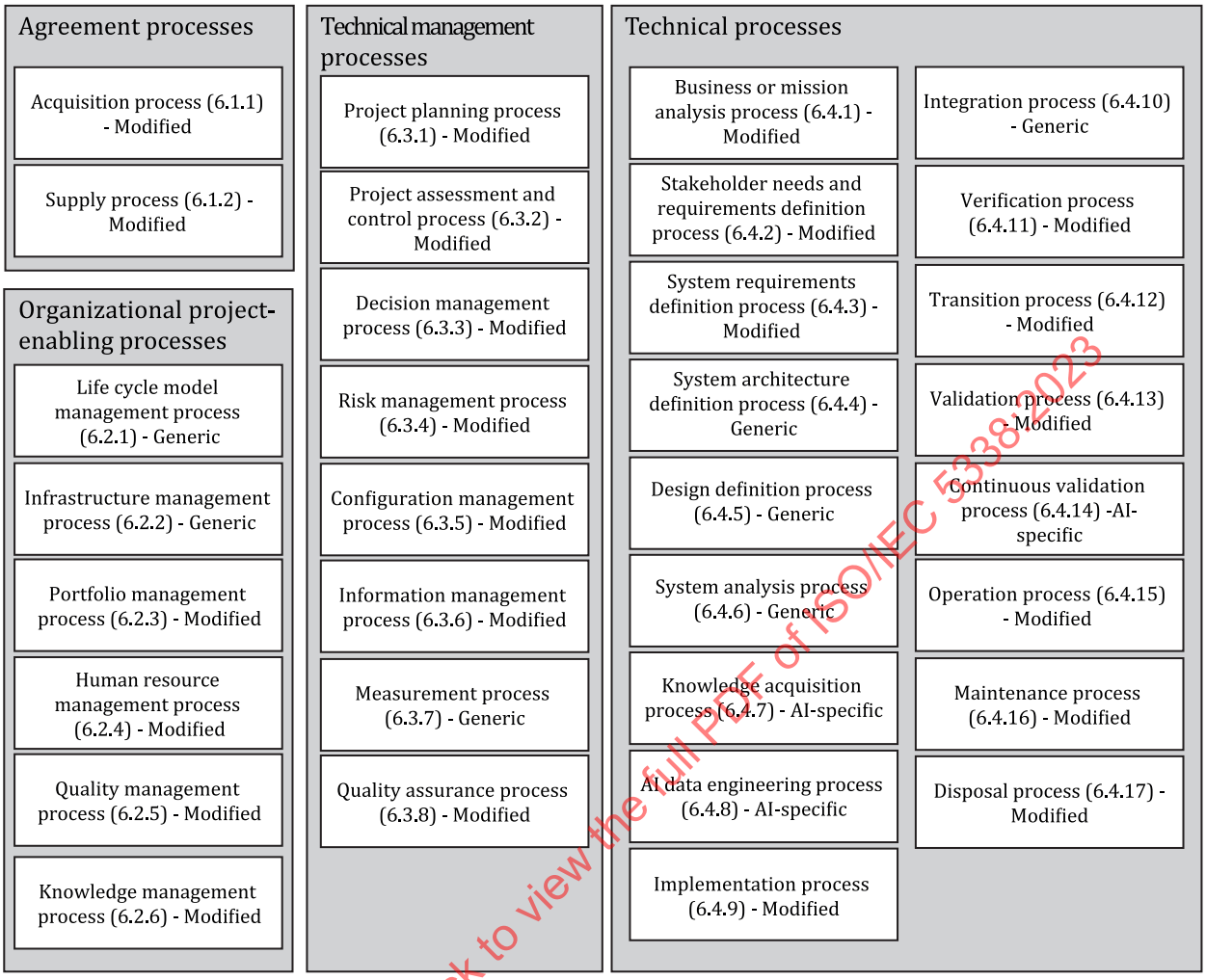

Figure 1 — AI system life cycle processes relative to ISO/IEC/IEEE 15288:2023, Figure 4

The following aspects of AI systems are key factors that differentiate the life cycle processes from those that are traditional systems.

— Measurable potential decay: Since AI models aim to model a desired behaviour which can change over time, measuring and monitoring any deviations of the production data (data drift) or deviations towards the desired output (concept drift) can be required. The changing of desired behaviour is not restricted to AI systems only, but for AI models this is uniquely measurable by validating input and output.

— Potentially autonomous: AI system's ability to make automated, complex and fast decisions creates the potential to replace actions or processes otherwise executed by humans. Consequently, AI systems can require extra attention to ensure fairness, security, safety, privacy, reliability, transparency and explainability, accountability, availability, integrity and maintainability. The more likely an AI system is able to do harm, the more important this extra attention becomes. See ISO/IEC TR 24368[14] for an overview of ethical and societal concerns in the development and deployment of AI systems. See ISO/IEC 23894[11] for more information of risk management of AI systems.

— Iterative requirements and behaviour specification: AI systems can be based on iterative and agile requirements specification, knowledge specification, behaviour modelling and usability design. AI system development can take place through cycles of requirements specification, prototype demonstration and requirements refinement. This aspect differs from traditional software applications based on fixed, well-defined requirements. Further, as AI systems are used, the

requirements can also evolve as unseen situations arise and refined requirements, specifications and gaps are identified.

— Probabilistic: Decisions made by AI systems based on machine learning are inherently probabilistic. Therefore, it is important for stakeholders to recognize that decisions made by AI systems are not always correct. Formally testing the correctness of models has inherent limitations and uncertainties when it comes to guarantees.

— Reliant on data: AI systems based on machine learning rely on sufficient, representative data to train, test and validate models. The behaviour of machine learning models is not programmed but is instead learned from the data. Because of this, it is important that particular consideration be given to the data (e.g. data quality) that are required for an AI system for training, testing, verification and validation.

— Knowledge intensive: For heuristic models, knowledge acquisition is of relatively high importance, since the knowledge is coded explicitly in the model and determines its correctness.

— Novel: New knowledge and skills can be required for organizations designing, developing or using AI systems. Other stakeholders, such as AI system users, can be unfamiliar with AI. This can cause trust and adoption challenges. The novelty of AI can cause overconfidence and enthusiasm without fully accounting for AI system risks. The perception that AI systems can eventually replace humans or demonstrate intelligence can also impact how stakeholders view AI systems.

— Incomprehensible: In case of heuristic models or machine learning, model behaviour is emergent in the sense that it is not explicitly programmed but is instead the indirect result of knowledge engineering or derived from the training data. Stakeholders can find AI systems to be less predictable, explainable, transparent, robust and understandable than explicitly programmed systems. This can reduce trust in AI systems.

NOTE 2 A high-level overview of AI ethical and societal concerns can be found in ISO/IEC TR 24368.[14] More information on addressing ethical concerns during system design can be found in IEEE 7000-2021[20].

5.2 AI system concepts

A model can be a machine learning model which has learned how to compute based on data, or it can be a heuristic model engineered based on human knowledge. In a heuristic model, the computations are engineered explicitly (procedural), implicitly by specifying rules or probabilities (declarative), or both.

In the case of machine learning, the data are the primary input for the model. For a heuristic model, the primary input is knowledge. Regardless, both data and knowledge are required in either case. Data are needed to test heuristic models and to perform analysis to build the knowledge. Knowledge is required to understand the context in which a machine learning model operates and to help select and prepare data for training and testing.

For traditional systems, both knowledge and data are often important as well. Knowledge can be required to implement business logic. Data typically plays an important part in any data processing system and can be required for functional testing.

The differentiation between an AI system and an AI application is provided in ISO/IEC 5339[3]. The distinguishing characteristics of AI applications are also defined in ISO/IEC 5339[3].

5.3 AI system life cycle model

The AI system life cycle model describes the evolution of an AI system from inception through retirement. This document does not prescribe a specific life cycle. Instead, it concentrates on AI-specific processes that can occur during the system life cycle. AI-specific processes can occur during one or more of the life cycle stages and individual stages of the life cycle can be repeated during the system's existence. For example, during the re-evaluation stage development and deployment can be repeated multiple times to implement bug fixes and system updates.

A system life cycle model helps stakeholders build AI systems more effectively and efficiently. International Standards are useful in developing the life cycle model, including ISO/IEC/IEEE 15288 for systems as a whole, ISO/IEC/IEEE 12207 for software and ISO/IEC/IEEE 15289[10] for system documentation. These International Standards describe life cycle processes for traditional systems. Figure 2 is based on ISO/IEC 22989:2022, Figure 3. It provides an example of the stages and high-level processes that can be applied to the development and life cycle of AI systems. For details, see ISO/IEC 22989:2022, 6.1.

The AI system life cycle or any subset of its stages can be owned and managed by separate organizations or entities (e.g. acquisition and provision of data, ML model or the code for other components used for the AI system development or deployment). Additionally, an organization can depend on other organizations to establish the infrastructure or to provide the necessary capability of the AI system life cycle (e.g. infrastructure setup cutting across on-premise, cloud-based or hybrid). This document takes into account the implications, specifics and associated risks of the AI system supply chain to propose new processes, adapt and tailor existing processes to build an AI system across organizational boundaries.

In addition, certain domains have specific life cycle International Standards such as medical devices, where IEC 62304:2006+A1:2015[19] applies. Organizations should consider the AI specifics described in this document together with IEC 62304:2006+A1:2015[19] when implementing such domain specific standards.

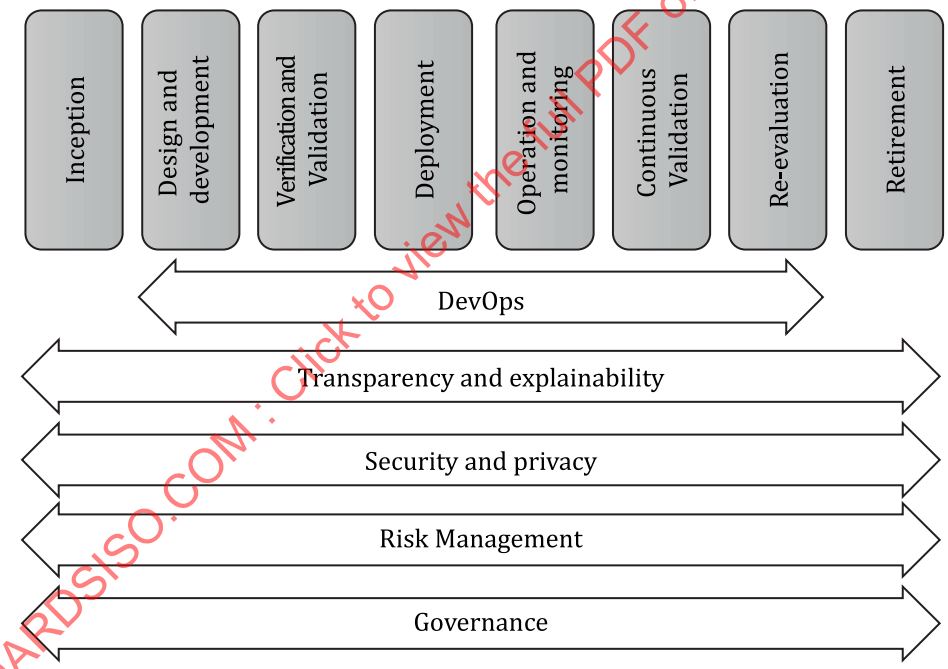

Figure 2 — Example of AI system life cycle model stages and high-level processes

The stages in Figure 3 are based on the stages described in ISO/IEC 22989 together with grouping of technical processes described in this document. The stage “continuous validation” is not marked as “in case of continuous learning”, in contrast with the example life cycle model in ISO/IEC 22989:2022, Figure 4. The continuous validation stage is also applicable in situations without continuous learning, for example, to detect data drift, concept drift or to detect any technical malfunctions.

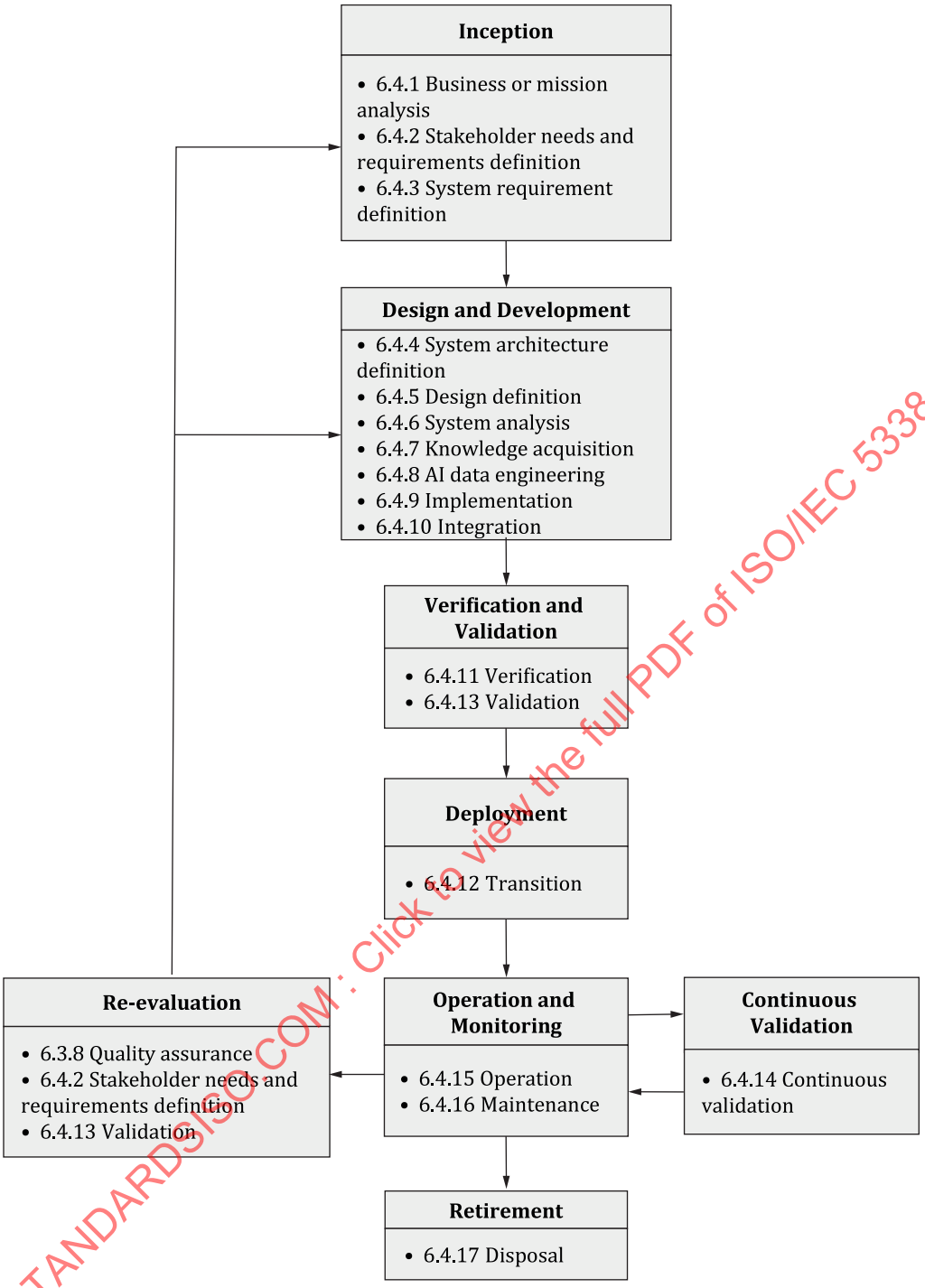

Figure 3 — AI system life cycle stages with technical processes

The concept of stages is meant to group activities that have a certain chronological order, to illustrate their dependency, but it does not suggest complete separation of activities in time or in the organization. For example, in agile software development, development and operation are distinct stages which are performed concurrently. Nevertheless, a piece of functionality first should be implemented before it can be verified and then deployed.

Furthermore, the sequence of stages can take place counter to the direction of arrows, for example when after the verification and validation stage it is decided to perform some activities as part of the design and development stage again.

NOTE 1 Stages in the life cycle in Figure 3 can have entry and exit criteria based on the specific requirements of the system in question (see ISO/IEC/IEEE 24748-1[15]).

An AI model can be either a machine learning model or a heuristic model.

The key technical processes for developing ML models are integrated into the life cycle processes as follows:

— system requirements definition process: set model requirements;

— AI data engineering process: acquire and update data;

— AI data engineering process: prepare data;

— implementation process and maintenance process: (re)train and tune model;

— verification process: test model before deployment;

— transition process: deploy model;

— continuous validation process: test model after deployment.

For heuristic models, the key steps are integrated as follows:

— system requirements definition process: set model requirements;

— knowledge acquisition process: acquire knowledge;

— implementation process and maintenance process: create and update model;

— verification process: test model before deployment;

— transition process : deploy model.

NOTE 2 The final decision whether to develop an AI system or a traditional system is the result of the inception stage, where requirements, risks, business and stakeholder needs are taken into account.

Annex A provides an analysis of the results of applying the traditional system life cycle processes to the use cases of AI systems from ISO/IEC TR 24030[13].

5.4 Process concepts

5.4.1 Criteria for processes

The life cycle processes in this document are based on the same principles of ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207. The processes in this document exhibit a strong relationship between their outcomes, activities and tasks. In addition, their description minimizes dependencies amongst the processes and ensures that a process can be executed by a single organization or across multiple organizations. This is particularly critical as AI systems can be developed across or require capability or support from a supply chain of organizations.

5.4.2 Description of processes

The purpose describes the goal of the process and is unchanged if the named process is from ISO/IEC/IEEE 15288 or ISO/IEC/IEEE 12207. The outcome describes the result of the successful implementation of the process. The activities and tasks describe the implements of the process in accordance with applicable organization policies and procedures. The AI-specific particularities for

processes from ISO/IEC/IEEE 15288 or ISO/IEC/IEEE 12207 are described in the subclause called “AI-specific particularities”.

5.4.3 Conformance

Conformance with this document is defined as implementing all of the processes, activities and tasks identified in this document. If a process, activity or a task is not relevant to an AI system, the absence of that process, activity or task shall be justified and documented. The requirements in ISO/IEC/IEEE 15288:2023, 4.2 and 4.3 and ISO/IEC/IEEE 12207:2017, 4.2 and 4.3 shall also apply.

6 AI System life cycle processes

6.1 Agreement processes

6.1.1 Acquisition process

6.1.1.1 Purpose

The purpose of the acquisition process is to obtain a product or service in accordance with the acquirer's requirements.

NOTE The acquirer refers to the stakeholder's role "AI customer" and the supplier refers to "AI producer" and "AI provider" as defined in ISO/IEC 22989.

6.1.1.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.1.1 and ISO/IEC/IEEE 12207:2017, 6.1.1 shall apply.

6.1.1.3 AI-specific particularities

The acquisition process described in ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207 should be extended beyond the acquisition of products or services to include the possible acquisition of data for the AI data engineering process (see 6.4.8). This new kind of acquisition activity can introduce new acquisition issues such as costs, dependencies, continuity, availability and issues with data rights, rules and legal requirements regarding the use of the acquired data. For example, contracting and acceptance of training data is an important issue because contracting and acceptance of datasets is very difficult to formalize. In addition, the acceptance activities can be followed by iterations of development or retraining activities in parallel of operations in order to make the accepted dataset remaining in line with the operational and business requirements.

6.1.2 Supply process

6.1.2.1 Purpose

The purpose of the supply process is to provide an acquirer with a product or service that meets agreed requirements in the agreement.

NOTE The acquirer refers to the stakeholder's role "AI customer" and the supplier refers to "AI producer" and "AI provider" as described in ISO/IEC 22989.

6.1.2.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.1.2 and ISO/IEC/IEEE 12207:2017, 6.1.2 shall apply.

6.1.2.3 AI-specific particularities

There are no additional activities or tasks defined in the supply process. When implementing the activities and tasks in 6.1.2.2, the supplier should consider the following AI-specific particularities to propose, negotiate and agree with the acquirer of the AI system.

— conduct of proof-of-concept to initiate AI system development before deployment;

— provision, collection or acquisition of sufficient datasets for machine learning;

— monitoring or intervention of AI system during the operation in case its performance varies depending on machine learning production data;

— analysis and improvement of the AI system to address any deviations from required performance.

6.2 Organizational project-enabling processes

6.2.1 Life cycle model management process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.1 and ISO/IEC/IEEE 12207:2017, 6.2.1 shall apply.

NOTE A typical life cycle model of AI systems is described in 5.3.

6.2.2 Infrastructure management process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.2 and ISO/IEC/IEEE 12207:2017, 6.2.2 shall apply.

6.2.3 Portfolio management process

6.2.3.1 Purpose

The purpose of the portfolio management process is to initiate and sustain necessary, sufficient and suitable projects in order to meet the strategic objectives of the organization. This process commits the investment of adequate organization funding and resources and sanctions the authorities needed to establish selected projects. Continued assessments are performed in this process to confirm that they justify, or can be redirected to justify, continued investment.

6.2.3.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.3 and ISO/IEC/IEEE 12207:2017, 6.2.3 shall apply.

6.2.3.3 AI-specific particularities

There are no additional activities or tasks defined in the portfolio management process. When implementing the activities and tasks in 6.2.3.2, organizations should consider the following AI-specific particularities:

— In the definition and authorization of projects, AI can provide a potential new capability or business opportunity to innovate through a new project.

— In the identification and allocation of resources to new projects, take into account that AI requires specific expertise (see 6.2.4).

— Especially when AI is a new capability in an organization, it can be beneficial to identify any multi-project aspects, so a typical approach can be achieved through the reuse of common AI system elements or platforms and the exchange of knowledge between projects.

— When evaluating projects in the portfolio, specific AI risks should be taken into account (see 6.3.4), as well as AI-specific aspects regarding project planning. For example, experimentation can require long periods for training of acceptable ML models.

6.2.4 Human resource management process

6.2.4.1 Purpose

The purpose of the human resource management process is to provide the organization with necessary human resources and to maintain their competencies, consistent with business needs.

This process provides a supply of skilled and experienced personnel qualified to perform life cycle processes to achieve organization, project and stakeholder objectives.

6.2.4.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.4 and ISO/IEC/IEEE 12207:2017, 6.2.4 shall apply.

6.2.4.3 AI-specific particularities

There are no additional activities or tasks defined in the human resource management process.

The use of AI techniques brings new AI stakeholder roles into the life cycle. For example, data scientists, data engineers play additional roles as AI developers in machine learning. Knowledge engineers play an additional role as AI developers in knowledge engineering. When implementing the activities and tasks in 6.2.4.2, organizations should consider the skills of these additional roles.

Additionally, organizations new to AI should review existing human resources and determine the appropriateness of their competencies.

See ISO/IEC 22989:2022, 5.17 for more details of AI stakeholder roles (e.g. AI developers, AI providers, data providers).

6.2.5 Quality management process

6.2.5.1 Purpose

The purpose of the quality management process is to ensure that products, services and implementations meet both relevant organizational and project quality objectives as well as meeting relevant customer requirements.

6.2.5.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.5 and ISO/IEC/IEEE 12207:2017, 6.2.5 shall apply.

6.2.5.3 AI-specific particularities

There are no additional activities or tasks defined in the quality management process. When implementing the activities and tasks in 6.2.5.2, organizations should consider the following AI-specific particularities.

Organizations should consider implementation of the AI-specific particularities laid down in their quality management processes, including but not limited to their policies, objectives and procedures. Quality assurance as part of the quality management process and the evaluation thereof can take a more prominent role in organizations developing, deploying and monitoring AI systems.

The continuous quality management activities support the systematic assessment of the performance of an AI system throughout its life cycle, including its continued quality once it has been deployed.

6.2.6 Knowledge management process

6.2.6.1 Purpose

The purpose of the knowledge management process is to create the capabilities and assets that enable the organization to exploit opportunities to reapply existing knowledge.

This encompasses knowledge, skills and knowledge assets, including system elements.

The knowledge that is used to create the AI models is discussed in 6.4.7 and 6.4.9.

6.2.6.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.2.6 and ISO/IEC/IEEE 12207:2017, 6.2.6 shall apply.

6.2.6.3 AI-specific particularities

There are no additional activities or tasks defined in the knowledge management process. When implementing the activities and tasks in 6.2.6.2, organizations should consider the following AI-specific particularities:

— Elements of an AI system (e.g. datasets, data preparation scripts) should be considered for knowledge management, just like any other system element.

— Experimentation is an important aspect in the implementation of an AI system. Documenting experiments is important to prevent having to repeat previous experiments in the future; either by the same stakeholder or by another stakeholder. Furthermore, artefacts documenting experimentation provides important insights and lessons learned that can be used for further improvement.

— See 6.2.4 for more details of the human resource regarding data science expertise.

— See 6.4.8 for more details of data lineage and data provenance.

6.3 Technical management processes

6.3.1 Project planning process

6.3.1.1 Purpose

The purpose of the project planning process is to produce and coordinate effective and workable plans.

This process:

— determines the scope of the project management and technical activities;

— identifies process outputs, tasks and deliverables;

— establishes schedules for conducting tasks, including achievement criteria;

— estimates the required resources to accomplish tasks.

This is an ongoing process that continues throughout a project, with regular revisions to plans.

6.3.1.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.1 and ISO/IEC/IEEE 12207:2017, 6.3.1 shall apply.

6.3.1.3 AI-specific particularities

There are no additional activities or tasks defined in the project planning process. When implementing the activities and tasks in 6.3.1.2, either projects or organizations, or both should consider the following AI-specific particularities.

In implementing the activity "plan project and technical management", it is important to allow some flexibility with regards to model creation (see ISO/IEC/IEEE 15288:2023, 6.3.1.3 and ISO/IEC/IEEE 12207:2017, 6.3.1.3). Predictability of software development is already challenging and for model creation, this is even more the case. Creating a model can require AI data engineering, such as data gathering, data labelling and data pre-processing (see 6.4.8). For a machine learning-based AI system, creating a model can require iterations of experiments and experimentation with different strategies and tactics to achieve the desired model performance and qualities. For a knowledge engineering-based AI system, creating a model can involve knowledge acquisition and elicitation efforts.

Furthermore, project planning should take into account the various other AI particularities of the processes involved, such as establishing continuous validation (see 6.4.14).

6.3.2 Project assessment and control process

6.3.2.1 Purpose

The purpose of the project assessment and control process is:

— to assess if the plans are aligned and feasible;

— to determine the status of the project, technical and process performance;

— to direct execution to help ensure that: the performance is according to plans and schedules, within projected budgets and satisfies technical objectives.

This process periodically and at major events evaluates the project's progress and achievements against requirements, plans and the overall business objectives. Information is provided for management to act upon when significant variances are detected. This process can include redirecting the project activities and tasks to correct deviations and variations from other technical management or technical processes. Redirection can include replanning as appropriate.

6.3.2.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.2 and ISO/IEC/IEEE 12207:2017, 6.3.2 shall apply.

6.3.2.3 AI-specific particularities

There are no additional activities or tasks defined in the project assessment and control process. When implementing the activities and tasks in 6.3.2.2, either projects or organizations, or both should consider the following AI-specific particularities.

In implementing the activity "plan for project assessment and control", intervals can be determined (as defined in the quality management process) whether to update the AI system or the model (see ISO/IEC/IEEE 15288:2023, 6.3.2.3 and ISO/IEC/IEEE 12207:2017, 6.3.2.3).

The implementation of an AI system is less predictable as a result of its iterative and experimental nature. For example, its progress cannot be reliably measured by counting the produced lines of code.

6.3.3 Decision management process

6.3.3.1 Purpose

The purpose of the decision management process is to provide a structured, analytical framework for objectively identifying, characterizing and evaluating a set of alternatives for a decision at any point in the life cycle and selecting the most beneficial course of action.

6.3.3.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.3 and ISO/IEC/IEEE 12207:2017, 6.3.3 shall apply.

6.3.3.3 AI-specific particularities

There are no additional activities or tasks defined in the decision management process. When implementing the activities and tasks in 6.3.3.2, either projects or organizations, or both should consider the following AI-specific particularities.

The use of AI adds more uncertainty and complexity to a system through the introduction of new types of decisions (see ISO/IEC 25059:2023[16] and ISO/IEC TS 25058 [6] for the details of the quality of AI system and machine learning models).

The new types of decisions include, but not limited to:

— decisions to retire the AI system when effective behaviour is no longer in line with the requirements;

— decision to "refactor" the AI system when the model training makes effective behaviour no longer in line with the requirements (i.e. the system is reset and a new trained model is created);

— decision to update the specification(s) and the contract(s) between the acquirer, the user or the supplier in order to reflect the learned behaviour;

— decision to update the documentation in order to reflect the learned behaviour (considering here that evolution remains in line with the requirements and the contract).

For example, the organization should determine how the quality of machine learning models is measured when implementing the activity "Analyse the decision information" (see ISO/IEC/IEEE 15288:2023, 6.3.3.3 and ISO/IEC/IEEE 12207:2017, 6.3.3.3). These aspects make it even more important to have a pre-defined decision-making process.

NOTE Once an organization has decided on its use of AI, this can assist governing bodies with identifying both the points or "gates" at which key governance questions can arise and can be addressed by the governing body (see ISO/IEC 38507:2022, 5.3[17]).

6.3.4 Risk management process

6.3.4.1 Purpose

The risk management process is an ongoing process that systematically and continuously identifies, analyses, treats and monitors risk throughout the life cycle of a system product or service. It can be applied to risks related to the acquisition, development, maintenance or operation of a system.

6.3.4.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.4 and ISO/IEC/IEEE 12207:2017, 6.3.4 shall apply.

6.3.4.3 AI-specific particularities

There are no additional activities or tasks defined in the risk management process. When implementing the activities and tasks in 6.3.4.2, either projects or organizations, or both should consider the following AI-specific particularities as well as should refer to ISO/IEC 23894[11] for the details of risk management of AI systems. The objectives for risk management as defined in ISO/IEC 23894[11] include fairness, privacy, reliability, transparency, explainability, accountability, availability, integrity and maintainability.

The risk assessment activities should include all risks associated with the AI system and the implementation of appropriate risk treatment measures through a risk treatment plan and relevant risk management records as defined in ISO/IEC 23894.[11] ISO/IEC 23894 provides risk management guidance to organizations which develop, produce, deploy and use products, systems and services that utilize AI. It is not intended for the specific risk management of products and services using AI for objectives such as safety and security. Therefore, organizations that apply AI in products and services with safety and security objectives should consider applicable risk management International Standards in addition to the AI specifics around processes as described in ISO/IEC 23894.[11] Considerations regarding the functional safety of AI systems can be found in ISO/IEC TR 5469.[5] For example, developers of AI systems considered as medical devices, should address risk management in line with International Standards, such as ISO 14971[9].

In addition to guidance provided in ISO/IEC/IEEE 15288:2023, 6.3.4 and ISO/IEC/IEEE 12207:2017, 6.3.4, AI systems have additional areas of opportunity or concern compared to traditional software systems. These areas are highlighted and explained in more detail in ISO/IEC 23894[11].

Another AI-specific consideration applies when AI systems are programmed to compute autonomous on-going decisions that involve risk of harm and where such decisions, due to time constraints, cannot be verified by a human being (e.g. some self-driving vehicles decisions). These risks can be mitigated by on-going risk management by the system itself. A simple form of this is setting specific boundaries in which the system can operate. For example, an automated climate control system should not allow heating above dangerous temperature ranges. Certain rules can also help to manage risks, such as do not open the trunk automatically at high speeds, even if the driver seems to have requested it. The most advanced type of on-going risk management is when an AI system actively performs a decision risk analysis through reasoning, based on a model of the world and rules. Apart from autonomous risk management, risk of harm can be mitigated by sufficient test case coverage, to verify that harmful decisions are not made in riskful situations. Additionally, for machine learning, extra care can be taken to include riskful situations and their correct decisions as cases in the training data as well. Human retrospective analysis of automated risk management can also be performed at a later stage to ensure quality.

See also 6.4.3 about the system requirements definition process for important properties of AI systems. In addition to the typical risks considered for a system, such as security and privacy, the risk treatment plan should also address risks related to the objectives as identified by the organization.

Organizations should identify potential risks and opportunities related to the AI system including conferring with representative users and other stakeholders to ascertain their needs and requirements (6.4.2).

Additional risk management requirements can apply based on the purpose of the AI system and the applicable regulatory environment in which the AI system is intended to be used.

If the AI system is related to safety, in order to establish accountability, the organization shall have an audit trail, including elements such as data provenance, data source validation, risk analysis and mitigation and decisions. This can be recommended to other AI systems, too. Therefore, development of an AI system shall include an audit strategy. For example, storing past decisions together with a reference to the model used including a trail on how that model was created. This can include documenting key decisions in the development process itself and their rationale (e.g. why a certain model was preferred).

6.3.5 Configuration management process

6.3.5.1 Purpose

The purpose of the configuration management (CM) process is to manage and control system elements and configurations over the life cycle. CM also manages consistency between a product and its associated configuration definition.

NOTE See ISO 10007[7] for more details of configuration management.

6.3.5.2 Outcomes, activities and tasks

The outcomes, activities or tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.5 and ISO/IEC/IEEE 12207:2017, 6.3.5 shall apply, with the following addition.

The project shall implement the following activities in accordance with applicable organization policies and procedures with respect to the configuration management process:

— An automated process for model rollback can be used to quickly resolve suboptimal model performance.

6.3.5.3 AI-specific particularities

There are no additional activities and tasks defined in the configuration management process. When implementing the activities and tasks in 6.3.5.2, either projects or organizations, or both should consider the following AI-specific particularities.

Apart from traditional software components and configuration, AI systems contain AI-specific artefacts that also require configuration management: the data that represents the model (e.g. rules, weights, parameters), documentation of AI elements, data and metadata. If machine learning is used, it can be beneficial to apply configuration management on the model combined with the data with which it was trained. This allows for traceability (e.g. for auditing and compliance) and for reproducing experiments.

Compared to the artefacts of traditional software (e.g. source code, test cases, test data), the artefacts of AI systems, especially datasets, can be large in size and are typically stored in systems separate from code and configuration files. In some cases of AI systems, the older data repository should be kept intact for the possible need of application version rollback. This can lead to choices that can compromise typical practice for traditional software, such as shorter retention periods of versions.

A typical application of configuration management for AI system is rollback to a previous version of a runtime model when a new model turns out to have quality issues. Information that is dealt with in a configuration management process includes the data that are used to build and test the AI model. More details specific to those data can be found in 6.4.7 and 6.4.8. Additionally, the data that are used to build and test the AI model are placed under configuration management where AI system performance in service is continuously monitored and maintained (see 6.4.14, 6.4.15 and 6.4.16).

Additionally, the version management of an AI system is no longer sufficient to provide a clear status of a configuration since the development and logistics versions of the configuration items no longer reflect the guaranteed behaviour associated to the operational configuration. In particular, if several instances of the same configuration are deployed, their behaviours can be different.

In the design and development stages, the organization should consider AI-specific source code management controls associated with AI-specific particularities (e.g. AI data engineering, model training).

6.3.6 Information management process

6.3.6.1 Purpose

The purpose of the information management process is to generate, obtain, confirm, transform, retain, retrieve, disseminate and dispose of information, to designated stakeholders.

An information management process plans, executes and controls the provision of unambiguous, complete, verifiable, consistent, modifiable, traceable and presentable information to designated stakeholders. Information includes technical, project, organizational, agreement and user information. Information is often derived from data records of the organization, system, process or project.

6.3.6.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.6 and ISO/IEC/IEEE 12207:2017, 6.3.6 shall apply.

6.3.6.3 AI-specific particularities

There are no additional activities or tasks defined in the information management process. When implementing the activities and tasks in 6.3.6.2, either projects or organizations, or both should consider the following AI-specific particularities.

AI systems are typically very data intensive and involve datasets for testing and in case of machine learning datasets for training. These datasets are part of the information that should be managed (see 6.4.8).

6.3.7 Measurement process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.7 and ISO/IEC/IEEE 12207:2017, 6.3.7 shall apply.

In addition, processes for AI-specific measurements shall be considered (e.g. probability of erroneous output) if the AI system is related to safety but they are recommended to other AI systems, too. Specifically, the drift in AI models due to environment changes and due to autonomous changes can be measured for corrections.

6.3.8 Quality assurance process

6.3.8.1 Purpose

The purpose of the quality assurance process is to help ensure the effective application of the organization's quality management process to the project.

Quality assurance focuses on providing confidence that quality requirements are fulfilled. Proactive analysis of the project life cycle processes and their outputs is performed to ensure that the product under development will be of the desired quality and that organization and project policies and procedures are followed.

6.3.8.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.3.8 and ISO/IEC/IEEE 12207:2017, 6.3.8 shall apply.

Quality assurance as part of the quality management process and the evaluation thereof can take a more prominent role in organizations developing, deploying and monitoring AI systems.

6.3.8.3 AI-specific particularities

There are no additional activities or tasks defined in the quality assurance process. When implementing the activities and tasks in 6.3.8.2, either projects or organizations, or both should consider the following AI-specific particularities.

In addition to the guidance provided in ISO/IEC/IEEE 15288:2023, 6.3.8 and ISO/IEC/IEEE 12207:2017, 6.3.8, quality assurance can take a more prominent role in AI systems compared to traditional software systems. AI systems can evolve over time, e.g. in continuous learning systems. Such evolution requires close quality monitoring and assurance activities to detect possible deteriorating effectiveness caused by, e.g. low quality of input data to the model, concept drift or data drift.

In the quality assurance process, both the product and the process are monitored and evaluated. In AI systems, algorithms and data for ML are also considered as products to be evaluated. In the evaluation of these products, the quality characteristics specific to AI systems (e.g. transparency, fairness, accountability, robustness) should be additionally included. Further information on the quality aspects of AI systems can be found in ISO/IEC 25059[16] and ISO/IEC TS 25058[6].

In addition, the processes to be evaluated should include activities for conducting analysis during the proof-of-concept, tasks for analysing requirements and risks to ensure adequate coverage of the problem domain of interest, iterative tasks for conducting machine learning, or procedures for creating training data (collection, selection, generation, validation and modification or addition).

More details specific to these data, processes and their quality evaluations for ML can be found in 6.4.7, 6.4.8 and 6.4.14.

Examples of effects that should be monitored by means of quality assurance include:

— data offered to the model is of low quality;

— data evaluated by the model is subject to change (data drift);

— deviations towards the desired output (concept drift).

Quality assurance activities should be appropriate to the use of the AI system. Typically, the complexity of the environment, the level of autonomy exercised, as well as the impact of the output of the AI system influences the level to which organizations should implement quality assurance activities. Moreover, there can be external factors such as regulatory requirements and quality system requirements that affect the type and extent of quality assurance activities. Especially for continuous learning AI systems, organizations should consider appropriate revalidation activities.

See 6.2.5, 6.4.11, 6.4.13, 6.4.14 and the description of data quality analysis in 6.4.8 for details.

6.4 Technical processes

6.4.1 Business or mission analysis process

6.4.1.1 Purpose

The purpose of the business or mission analysis process is to define the business or mission problem or opportunity, characterize the solution space and determine potential solution class or solution classes that can address a problem or take advantage of an opportunity.

6.4.1.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.1 and ISO/IEC/IEEE 12207:2017, 6.4.1 shall apply.

6.4.1.3 AI-specific particularities

There are no additional activities or tasks defined in the business or mission analysis process. When implementing the activities and tasks in 6.4.1.2, either projects or organizations, or both should consider the following AI-specific particularities.

AI systems involve specific risks (see 6.3.4) that can influence or even prohibit the realization of certain business goals. For example, when implementing the "Define the problem or opportunity space" activities, the organization shall consider that privacy regulations can prevent the use of personal data for a use case that differs from its original collection intent (see ISO/IEC/IEEE 15288:2023, 6.4.1.3 and ISO/IEC/IEEE 12207:2017, 6.4.1.3). In case of making decisions that can affect individuals negatively, legal requirements can require explanation of what data were used and how data were used.

Other examples of risks are the degree of data accessibility and data quality.

6.4.2 Stakeholder needs and requirements definition process

6.4.2.1 Purpose

The purpose of the stakeholder needs and requirements definition process is to capture the stakeholder requirements for a system so it can provide the capabilities needed by users and other stakeholders in a defined environment.

It identifies stakeholders, stakeholder classes and their needs throughout the life cycle of the AI system. It analyses and transforms these needs into a common set of stakeholder requirements that express the intended interaction the system will have with its operational environment and that are the reference against which each resulting operational capability is validated. The stakeholder requirements are defined considering the context of the system-of-interest with the interoperating systems and enabling systems.

6.4.2.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.2 and ISO/IEC/IEEE 12207:2017, 6.4.2 shall apply.

6.4.2.3 AI-specific particularities

There are no additional activities or tasks defined in the stakeholder needs and requirements definition process. When implementing the activities and tasks in 6.4.2.2, either projects or organizations, or both should consider the following AI-specific particularities.

Due to the inductive nature of AI systems, it is crucial to execute the task to "obtain explicit agreement on the stakeholder requirements"; including critical performance measures that enables the assessment of technical achievement as a target (see ISO/IEC/IEEE 15288:2023, 6.4.2.3 and ISO/IEC/IEEE 12207:2017, 6.4.2.3). Organizations should consider the possibility of introduction of biases by narrow views of stakeholders.

Such a technical achievement should be specified by the organization to allow for monitoring of the targets through the quality assurance process (see 6.3.8).

The use of the AI system can define particular stakeholder types that should be considered. Particular types of stakeholders to consider include:

— providers of AI platforms, products or services;

— AI developers;

— customers and users;

— partners involved in system integration, data provision and auditing;

— regulatory and policy-making authorities, data subjects;

— others impacted by the development and use of the AI system.

The stakeholder identification process can yield a set of values that can guide development of parts of the system including features such as the user interface, documentation and use cases. Organizations should further study and refine the values to the extent they can become part of the system requirements. Regulation, human rights, social responsibilities and environmental frameworks can help in refining and describing the values. See ISO/IEC 22989:2022, 5.17 for more details of possible AI stakeholder types.

NOTE The quality characteristics of the quality model of AI systems that are available in ISO/IEC 25059[16] are useful to elicit and identify quality requirements of non-functional requirements, which are often implicit stakeholder needs. Also see ISO/IEC TS 25058[6] for more information on the quality evaluation of AI systems.

6.4.3 System requirements definition process

6.4.3.1 Purpose

The purpose of the system requirements definition process is to consider and transform all stakeholder requirements into a technical view of a solution that still meets the operational needs of the user. The process factors in outputs from the risk management and governance process in particular, as indicated by Figure 2.

This process creates a set of measurable system requirements that specify, from the supplier's (who has a role of AI producer, AI partner or AI provider) perspective, the system's characteristics, attributes, functional and performance requirements, in order to satisfy the stakeholder requirements. The requirements should not imply any specific implementation.

6.4.3.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.3 and ISO/IEC/IEEE 12207:2017, 6.4.3 shall apply.

6.4.3.3 AI-specific particularities

There are no additional activities or tasks defined in the system requirements definition process. When implementing the activities and tasks in 6.4.3.2, either projects or organizations, or both should consider the following for AI systems:

— The desired performance (level of correctness) of the model or models. Setting these requirements requires careful selection of the right metrics (e.g. minimal precision and minimal accuracy). Such requirements can include the range of input data for which the model is to perform within the required boundaries. For example, the model is able to distinguish a cat from a dog in 90 % of the cases from photographs taken during daytime where the entire animal is visible.

— Requirements on the level of autonomy exercised by the AI system. These include considerations regarding the level of autonomy exercised by the AI system, e.g. whether there is a human-in-the-loop. If so, a definition of which decisions the human can take with regards to the AI system behaviour, such as setting or adjusting thresholds configuring the desired performance level of the AI system.

— Requirements on how to deal with unexpected behaviour of the system. For example, by establishing and applying additional deterministic rules to ensure safety.

— System performance requirements: such as the desired execution time, which often depends on the type of model used, should be defined.

— Requirements on transparency and explainability. Machine learning models can be highly complex and, as a result difficult to understand. Depending on the situation, individuals can be entitled to an

explanation as to how a model decision was made particularly when they are affected significantly (e.g. legally or financially). For example, in some countries an explanation is necessary when a credit application is denied. Explanations can vary from detailed to a high-level description of what data and what type of machine learning algorithm were used. Explanations can help acceptance of AI decisions but also to raise issues when the explanation indicates an error.

— The organization informs individuals that they are interacting with an AI system in accordance with applicable legal requirements.

— Continuous validation requirements: See the continuous validation process (6.4.14).

— Fairness requirements: It is important to set requirements for the fairness and inclusiveness of the algorithm and data for certain groups in society. Furthermore, AI system decisions should be based on clear and interpretable features so that fairness can be verified. Fairness metrics should be defined in order to set these requirements.

— Privacy requirements: In cases where personal data are used. Informing individuals, providing them with control and protection of personal data are important. Also, the choice of algorithm can be influenced by privacy considerations (e.g. differentially privacy algorithms[23]).

— Security requirements: In case there is an additional attack surface resulting from the use of AI. Typically, this includes:

  — securing data that are used for either training or testing, or both, including protection against "poisoning attacks" when malicious actors inject data to influence the behaviour of machine learning models;

  — protecting against input manipulation (e.g. a spam email being classified as not spam);

  — protecting against "model inversion" when a malicious actor manages to deconstruct sensitive data that are used for training a model;

  — protecting against "model theft" when a malicious actor aims to copy the behaviour of a model that is intellectual property.

6.4.4 System architecture definition process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.4 and ISO/IEC/IEEE 12207:2017, 6.4.4 shall apply.

NOTE The system architecture definition process is named as "architecture definition process" in ISO/IEC/IEEE 12207:2017, 6.4.4.

6.4.5 Design definition process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.5 and ISO/IEC/IEEE 12207:2017, 6.4.5 shall apply.

6.4.6 System analysis process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.6 and ISO/IEC/IEEE 12207:2017, 6.4.6 shall apply.

6.4.7 Knowledge acquisition process

6.4.7.1 Purpose

The purpose of the knowledge acquisition process is to provide the knowledge necessary to create the AI models.

Knowledge about the domain and the problem is central to many AI systems.

For a machine learning-based AI system, knowledge is used to guide the tasks of data selection, data preparation and model engineering. A knowledge acquisition process can be performed either by doing research or by involving subject matter experts, or both.

For a knowledge-based AI system, the knowledge should be coded explicitly in the model.

Data analysis (see 6.4.8) can play a part in gathering and refining knowledge.

NOTE Knowledge in the knowledge acquisition process is the knowledge necessary to create the AI models.

6.4.7.2 Outcomes

As a result of the successful performance of the knowledge acquisition process:

a) Knowledge necessary to create the AI models is identified.

b) Gathered knowledge is stored.

c) Traceability of knowledge acquisition is established.

6.4.7.3 Activities and tasks

The project shall implement the following activities in accordance with applicable organization policies and procedures with respect to the knowledge acquisition process:

a) Define scope and criteria for knowledge acquisition. As a first step, the scope and criteria of the knowledge are defined: What domain and what aspect is the knowledge about? How recent is the knowledge?

b) Seek and gather sources of knowledge. Knowledge can be extracted from publications, data or experts.

c) Perform knowledge acquisition to extract knowledge. In order to utilize the knowledge, publications can be studied, data can be analysed and experts interviewed or observed. In case of knowledge engineering, the extracted knowledge should be formalized in such a way that the algorithms involved can utilize it. This is part of the implementation process (see 6.4.9).

d) Knowledge about the domain and the problem is gathered through studying, interviewing or other knowledge elicitation, data analysis, acquiring documented knowledge or involving stakeholders with the required knowledge.

e) Manage the result of knowledge acquisition.

NOTE 1 The roles, activities, constructional layers, components of knowledge engineering and their relationships, as well as a common knowledge engineering vocabulary are provided in ISO/IEC 5392[4].

NOTE 2 Knowledge collected for a project can also be shared between projects via repositories (areas storing knowledge sets) and registries (entities recording the usages of knowledge sets). Reuse of knowledge can be considered during the knowledge collection process with regards to proven patterns which can apply in the model development activities (see 6.4.9.3).

6.4.8 AI data engineering process

6.4.8.1 Purpose

The purpose of AI data engineering process is to make sure data can be used to create AI models and verify AI models. Data are central in engineering machine learning models, as it is used to train them. For heuristic models, the purpose of data in model creation is more secondary, as it can be used to support knowledge engineering (see 6.4.9).

6.4.8.2 Outcomes

As a result of the successful performance of the AI data engineering process:

a) Required data and datasets are identified, sampled and obtained.

b) Training data and, if necessary, validation data are prepared, formatted and made available to machine learning models.

c) Test data is prepared for testing or validation (see 6.4.11).

d) Data for manual analysis to gain more understanding in order to support the AI data engineering and model engineering processes are prepared.

e) Automated processes, if any, to extract, transform and load the data are identified.

f) Any recording and use of personal information in the data are in compliance with applicable laws and legal requirements.

g) Artefacts (e.g. metadata) for traceability, documentation, maintenance of the data and the automated process, including configuration management are prepared.

h) Data are retired in a timely manner.

i) Multi-modal data are managed.

NOTE Because the use of multi-modal type of data (e.g. voice, images, sensors) within AI systems is growing, best practices can be employed to handle, engineer and deploy multi-modal AI systems.

6.4.8.3 Activities and tasks

The project shall implement the following activities in accordance with applicable organization policies and procedures with respect to the AI data engineering process.

a) Acquire or select data

The purpose of an AI model is to create an output based on an input, for example, classifying an animal based on an input image and therefore data should be gathered to constitute such input-output combinations. Typical forms of data are structured data, text, sound, image and other sensory data.

Examples of ways to gather data are:

— collecting from an existing data store (e.g. customer data);

— recording from a process (e.g. from industrial sensors);

— recording from an orchestrated process (e.g. acting scenes to create video examples for detecting certain events).

In order to test the ability of a machine learning model to generalize beyond the training data, it is helpful for test data to be from a different source or process. A famous example is where a machine learning model learned to recognize wolves from labelled training data. It turned out the model was able to perform well because all wolf training pictures were taken in winter and can be identified easily by detecting snow. In order to prevent such generalization issues, test data should have been collected from a different source.

Data acquisition and selection should be an ongoing or recurring process in situations where the relation between the input and output variables changes over time. For example, in order to predict the selling price for a piece of land, it is important to keep up to date with changes in the economy and the market, which are reflected in new data. This new data can be used to test the model regularly and

when necessary to retrain, or re-engineer the model (see 6.4.14). Older data that represents outdated relations should be retired for the same reasons.

b) Conduct data labelling

Data labelling is a special form of data acquisition in which cases are assigned the value of the desired output, e.g. labelling images of animals with either “cat” or “dog”. This is typically done manually and therefore a strictly controlled process can help to prevent unwanted bias or noise from subjective elements.

Persons performing the labelling should be competent in the domain of what is being labelled and trained on the use of the labelling tool. Depending on the risk of the application, labelling results can be subject to review and correction, if necessary.

When making use of tools to support the labelling of data, organizations should consider the status of the tools used for the labelling process. Such considerations should include an evaluation of the features and functionalities of such annotation tools and the proper validation of such tools to ensure the high quality of labelled data (reference is made to 6.4.8.3, e) and f)).

c) Analyse and explore data for understanding

Gathered data can be analysed and explored to help the understanding of the domain, the problem and data issues. For machine learning, this understanding can lead to new insights into either what other data are needed or what processing of the data is required, or both. For knowledge engineering, the data analysis can be beneficial in further shaping existing knowledge.

d) Analyse data quality

Data can have many quality issues, which require assessment to guide selection, cleaning and correction of data. Data should be of sufficient quantity and bias should be within acceptable limits. It should be sufficiently complete (broad and varied) to be representative of the expected production data. Training data preferably has the same balance (distribution) as the model can expect to see, yet it is necessary to consider specific edge cases.

Bias (e.g. from subjective decisions) can be verified by checking if desired behaviour is well balanced with respect to social attributes for which discrimination is prohibited (e.g. gender or ethnicity). Removing such attributes is often not sufficient to deal with bias, as the input data can still have data elements that are proxies to those attributes.

A special aspect of data quality is the risk of the data being poisoned: A malicious actor can change, remove or add data to influence the model behaviours in an undesired way.

Data quality analysis is typically an ongoing process, as new quality issues can occur over time, which is why it is advisable to automate quality checks and verification. Such automation also serves as documentation of what checks are required.

NOTE Additional information regarding data quality is available in ISO/IEC 5259-1,[2] ISO/IEC 5259-2,[24] ISO/IEC 5259-3,[25] ISO/IEC 5259-4,[26] on data quality for analytics and ML. Additional information regarding different forms of bias in data used in AI systems is available in ISO/IEC TR 24027[12].

e) Document data lineage and data provenance

Since training data can determine the behaviour of an AI system, it is important to understand its source, how it was processed, its owner and its rationale, in case there are any issues with the data or its need to be renewed.

Metadata on data lineage documents the data origin, what happens to it and where it moves over time. Data lineage gives visibility while greatly simplifying the ability to trace errors back to the root cause in a data analytics process as well as to trace errors forward into product versions when a problem has been found with the source data.

Metadata on data provenance documents the inputs, entities, systems and processes of the data of interest, in effect providing a historical record of the data and its origins.

See 6.2.6 for more information on knowledge management and 6.3.6 on information management.

f) Clean, merge and prepare data

Data preparation is the collection of operations on data that lead to the desired outcome, including extraction, merging, cleaning, filtering, correcting, augmenting, conversion, encoding and dealing with missing values.

The goal of data preparation is to produce features for the data that are used as input for the AI model. Feature engineering is the process of selecting, characterizing and optimizing features for use in an AI model. In this process, selecting appropriate input can be done using domain knowledge, data analysis or experimenting with different selections of features. Some model types include optimising the selection of features. In general, the fewer features that are used, the easier it is to train a machine learning model and the fewer risks of data errors and the less effort in AI data engineering.

Filtering removes unwanted data that is:

— not useful to create or verify the model (e.g. outliers in some situation);

— unnecessary in volume, so a sample can suffice;

— harmful because it introduces unwanted bias or unwanted discrimination (e.g. related to gender or ethnicity);

— against privacy legal requirements, requiring removal or de-identification of personal data;

— sensitive data that should be protected from unauthorized internal or external access.

In some situations, data augmentation can help to increase the volume of data to create a better model or to perform more testing (e.g. rotating images).

Data conversion and feature encoding are used to transform data so as to satisfy the criteria that the AI model has towards input data (e.g. strictly yes or no variables).

Generative AI methods can be adopted to automatically generate metadata to support AI model by identifying patterns in production data.

Data preparation is often an exploratory and therefore less structured process, consisting of manual steps or code that is created ad hoc. This ad hoc nature can make repeated preparation more difficult and therefore it is beneficial to aim for reusable automated preparation.

Given the typical high complexity and exploratory character of the data preparation, (automated) testing of the preparation is important.

g) Protect sensitive data

Some aspects of AI systems rely on sensitive data and when it does, the AI data engineering process increases the attack surface of an AI system. This means that apart from the AI system itself, elements of AI data engineering can be subjected to an attack, such as a data store. Especially when data about individuals are gathered from different sources, security and privacy risks occur. In such cases, careful handling and privacy-preserving techniques are necessary.

6.4.9 Implementation process

6.4.9.1 Purpose

The purpose of the implementation process is to realize a specified system element.

This process transforms requirements, architecture and design – including interfaces – into actions that create a system element according to the practices of the selected implementation technology, using appropriate technical specialties or disciplines. This process results in a system element that satisfies specified system requirements (including allocated and derived requirements), architecture and design.

6.4.9.2 Outcomes

The outcomes provided in ISO/IEC/IEEE 15288:2023, 6.4.7 and ISO/IEC/IEEE 12207:2017, 6.4.7 shall apply.

As a result of the successful implementation of the AI model engineering part of the implementation process:

a) A working AI model is produced.

b) Documentation of the modelling process is produced.

6.4.9.3 Activities and tasks

The activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.7 and ISO/IEC/IEEE 12207:2017, 6.4.7 shall apply, with the following addition.

The project shall implement the following activities in accordance with applicable organization policies and procedures with respect to the AI model engineering part of the implementation process.

For machine learning-based AI systems, the following additional activities are included:

a) Algorithm selection: Selecting an appropriate machine learning algorithm taking into account the type of model task (e.g. clustering, time series prediction, classification) and the technique that works best for the task at hand, which can also be determined by experimentation.

   One aspect to take into account when selecting (and tuning) algorithms is how interpretable or explainable the model can be. Typically, the more difficult to interpret models are the best performing models. On the other hand, interpretable models help build trust and transparency. This transparency can be beneficial for accountability and assists AI developers to better understand the problem domain and the data.

b) Model training: Run the algorithm iteratively over the training data to establish an internal representation (e.g. weights in a neural network). For supervised learning, the goal is to use the examples in the training data for estimating the underlying function that maps input to the desired output (e.g. the classification of cat or dog based on an image). It is important for the model to generalize well on these examples by preventing overfitting, which is what happens when a model performs well on the training data and poorly on production data.

c) Model tuning: Employing optimization techniques to find the hyperparameters that provide the best performance, using validation data.

For knowledge engineering-based AI systems, the following additional activities are included:

d) Knowledge programming: After knowledge has been acquired (see 6.4.7) it should be formalized in a heuristic model where the computations are either engineered explicitly (procedural – more according to traditional software programming) or implicitly by either specifying rules or probabilities (declarative), or both.

Define and prescribe the combined architecture considering cloud and edge computing to manage the emerging behaviour of AI systems especially in industrial applications.

6.4.9.4 AI-specific particularities

When implementing the activities and tasks of this process, organizations should consider the following AI-specific particularities.

AI systems can be seen as traditional software systems that apply one or more AI models and therefore their implementation involves the same practices, with some particularities that also introduce new elements. An example is the general best-practice of working with an actively maintained list of agreed-upon work items (backlog). Involving the AI work in such a backlog facilitates cross-disciplinary coordination, planning and evaluation.

The AI model is typically part of an application that, apart from the model, is developed without any machine learning or knowledge engineering. Because data and knowledge each have very specific roles in AI, this document introduces AI model engineering as a part of the implementation process and also introduces a separate process for AI data engineering. Data and model engineering are closely related and many implementation activities combine both parts, yet they are different in nature.

In the case of machine learning, AI model engineering requires training a model using training data. This is an iterative optimization in which the type of model is selected, its hyperparameters are configured and changed until the model performs adequately on the training dataset. Therefore, the AI data engineering process typically interacts with AI model engineering and often strongly relies on experience of the experts involved. Automated machine learning (AutoML) is an approach that the whole or part of these processes are automated, to reduce this reliance and to make the work more efficient. Efficiency can also be gained by, where possible, distributing the work between experts and computer resources to experiment in parallel. Depending on the algorithms used and the application of AutoML, model training and other optimizations possibly need substantial computing power and waiting times.

When a model performs in line with either exceptions or predefined specifications, or both, it can be further tuned using validation data and then tested using test data (see 6.4.11).

A special type of model engineering is transfer learning, in which an existing machine learning model is used as a starting point to further train for a slightly different use case. By building on previous model engineering success, efficiency gains can be made.

For implementation, existing software frameworks can be built upon. These frameworks typically offer various built-in AI models and solutions for data processing, model training, testing and orchestration.

In the case of knowledge engineering, model engineering is the specification of knowledge in declarative or procedural form. The knowledge is acquired from either experts (knowledge elicitation) or through data analysis (see 6.4.8), or both.

6.4.10 Integration process

The purpose, outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.8 and ISO/IEC/IEEE 12207:2017, 6.4.8 shall apply.

6.4.11 Verification process

6.4.11.1 Purpose

The purpose of the verification process is to provide objective evidence that a system or system element fulfils its specified requirements and characteristics.

The verification process identifies the anomalies (errors, defects or faults) in any information item (e.g. system requirements or architecture description), implemented system elements, or life cycle processes using appropriate methods, techniques, standards or rules. This process provides the necessary information to determine resolution of identified anomalies.

6.4.11.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.9 and ISO/IEC/IEEE 12207:2017, 6.4.9 shall apply.

6.4.11.3 AI-specific particularities

There are no additional activities defined in the verification process. When implementing the activities and tasks in 6.4.11.2, this process should be extended beyond the verification of systems to include the following AI-specific particularities.

The first AI-specific particularity that should be considered in the verification process is the verification through behaviour.

Whereas traditional systems are programmed to behave exactly as defined, AI models are constructed to come as close as possible to the desired behaviour. This probabilistic nature means these models should be verified statistically.

As AI models transfer an input to an output, verification of models typically happens by using verification datasets containing inputs and desired output and applying the statistical techniques to measure the desired correctness and robustness (see the system requirements definition process in 6.4.3). The datasets for verification can be collected from a separate and distinct subset of the same source of the training data or from a different source. The advantage of the latter is that a different source of data better tests the generalisation ability of the model.

Two types of the datasets for verification can be distinguished. Validation data is used to select the best model among candidate models. Test data is used to determine whether the final selected model performs and generalizes adequately.

The second AI-specific particularity requiring consideration in the verification process is the verification through review.

Code review as a verification method is suitable when it comes to source code specifically written for the AI system, including knowledge that is represented in heuristic systems. In the case of machine learning models, however, the source code of the algorithm cannot be reviewed when it is part of an existing library or framework. The behaviour of the machine learning model is determined by its representation in parameters. Even if the representation of the model is readable, its correctness is typically very difficult to assess, because the algorithm does not follow steps that have been implemented by programmers, following a relatable human thinking process. Instead, the algorithm follows steps that have been optimized in an automated way, to maximize the model's performance.

The organization should make sure that source code involved with AI is included in regular reviews of code quality aspects, such as maintainability, testability and reusability (e.g. peer reviews of training scripts or unit testing of data preparation code similar to source code that is not involved with AI).

See 6.4.14 for more details of the continuous validation process.

6.4.12 Transition process

6.4.12.1 Purpose

The purpose of the transition process is to establish a capability for a system to provide the services as specified by stakeholder requirements in the operational environment.

This process moves the system in an orderly and planned manner into its operational status, such that it is functional, operable and compatible with other operational systems. The process installs a verified system, together with relevant enabling systems, for example, planning system, support system, support staff, training system, user training system, as defined in agreements. The transition process is used at each level in the system structure and in each stage to complete the criteria established for exiting the stage. It includes preparing applicable storage, handling and shipping enabling systems.

6.4.12.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.10 and ISO/IEC/IEEE 12207:2017, 6.4.10 shall apply.

6.4.12.3 AI-specific particularities

There are no additional activities and tasks defined in the transition process. When implementing the activities and tasks in 6.4.12.2, either projects or organizations, or both should consider the following AI-specific particularities.

There is often a difference between the actual AI system that runs the model and the model itself – where the latter is a configuration or a set of parameters (e.g. the weights of the neural networks).

Because of operational requirements, AI models can be deployed in a different format from how they were developed.

The organization should aim to support model updates (retraining or knowledge engineering) and the execution of continuous monitoring of established metrics associated with the use of the AI system.

The organization should assess how performance of the AI system can be affected after it has been put into use and, by allowing for such factors, design appropriate monitoring metrics. An AI system, once deployed, can exhibit unexpected behaviour (e.g. as a result of bias or being exposed to unexpected input). Therefore, monitoring performance is important, and procedures and processes can be more extensive compared to traditional systems.

The possible timing of a model update is influenced by:

— changes of relevant operational processes;

— observation of changes in relevant data over time (e.g. data drift as explained in ISO/IEC 23053);

— observation of deterioration in precision (e.g. due to concept drift as explained in ISO/IEC 23053);

— the amount of time that has passed since the model creation of the last model update or model creation.

As some AI systems have the ability to improve their performance over time, the organization should support the attainment of quality of such improvements by implementing a quality management process. For example, monitoring trends of the amount of use associated with the AI system can support the organization ensuring its continued “quality in use”.

6.4.13 Validation process

6.4.13.1 Purpose

The purpose of the validation process is to provide objective evidence that the system, when in use, fulfils its business or mission objectives and stakeholder requirements, and achieves its intended use in its intended operational environment.

The objective of validating a system or system element is to acquire confidence in its ability to achieve its intended mission (or use) under specific operational conditions. Validation is ratified by stakeholders. The validation process provides the necessary information so that identified anomalies can be resolved by the appropriate technical process where the anomaly was created.

6.4.13.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.11 and ISO/IEC/IEEE 12207:2017, 6.4.11 shall apply.

6.4.13.3 AI-specific particularities

There are no additional activities defined in the validation process. When implementing the activities and tasks in 6.4.13.2, this process should be extended beyond the validation of systems to include the following AI-specific particularities.

In order to gain experience with AI systems, organizations can perform a proof-of-concept project. In such a case, validation also includes the validation of AI itself and its value and risks for the organization in general.

The organization should aim to support model updates (retraining or knowledge engineering) and the execution of continuous monitoring of established metrics associated with the use of the AI system.

The organization should assess how performance of the AI system can be affected after it has been put into use and, by allowing for such factors, design appropriate monitoring metrics. An AI system, once deployed, can exhibit unexpected behaviour (e.g. as a result of bias or being exposed to unexpected input). Therefore, monitoring performance is important, and procedures and processes can be more extensive compared to traditional systems.

The possible timing of a model update is influenced by:

— changes of relevant operational processes;

— observation of changes in relevant data over time (e.g. data drift as explained in ISO/IEC 23053);

— observation of deterioration in precision (e.g. due to concept drift as explained in ISO/IEC 23053);

— the amount of time that has passed since the model creation of the last model update or model creation.

As some AI systems have the ability to improve their performance over time, the organization should support the attainment of quality of such improvements by implementing a quality management process. For example, monitoring trends of the amount of use associated with the AI system can support the organization ensuring its continued "quality in use".

6.4.14 Continuous validation process

6.4.14.1 Purpose

The purpose of the continuous validation process is to monitor that AI models keep performing satisfactorily, or to demonstrate performance of the AI model over time.

AI models aim to model a desired behaviour and this desired behaviour can change. Also, production data can change over time. Therefore, it is important to measure and monitor deviations of the input data (data drift) or deviations towards the desired output (concept drift) using test data. This process serves as an extension of the quality assurance process (see 6.3.8).

If deviations are substantial, a machine learning requires retraining or continuous learning, as part of the maintenance process (see 6.4.16). Deviations can also point to other issues, for example data quality problems, or a system malfunctioning. If the AI system applies automated continuous learning without human interaction, an automated rollback process, at defined thresholds, should be included to prevent undesired model changes.

6.4.14.2 Outcomes

As a result of the successful performance of the continuous validation process:

a) Validation results are recorded in a validation log.

b) A decision can be made to perform maintenance on the AI model (retraining).

6.4.14.3 Activities and tasks

The project shall implement the following activities in accordance with applicable organization policies and procedures with respect to the continuous validation process:

a) Monitor for data drift by applying checks on the model input data to see if it deviates from what the model was trained with.

b) Monitor for concept drift by measuring model performance using test data that has been updated, or by detecting any anomalies in the output value or the distribution of output values, for example, by comparing recent output to previous output.

c) Monitor any other requirements that are expected to change over time (see 6.4.3) such as execution time, transparency and fairness.

d) In case of deviations, decide whether to perform maintenance on the AI model.

e) Apply guard rails if they have been defined by applying boundaries on the output data, or by defaulting to an alternative safe model in case of deviations.

f) Determine the frequency at which validation should occur.

6.4.15 Operation process

6.4.15.1 Purpose

The purpose of the operation process is to use the system to deliver its services.

This process establishes requirements for and allocates the personnel to operate the system. It monitors the services and evaluates operator-system performance. In order to sustain the operational services, it identifies and analyses operational anomalies in relation to agreements, stakeholder requirements and organizational constraints.

6.4.15.2 Outcomes, activities and tasks

The outcomes, activities and tasks provided in ISO/IEC/IEEE 15288:2023, 6.4.12 and ISO/IEC/IEEE 12207:2017, 6.4.12 shall apply.

6.4.15.3 AI-specific particularities

There are no additional activities defined in the operation process. When implementing the activities and tasks in 6.4.15.2, this process should be extended beyond the operation of systems to include the following AI-specific particularities.

The first AI-specific particularity that is considered in the operation process is computer resource and power usage.

AI systems can consume considerable computing power and memory usage, particularly the training of machine learning models (depending on the type of algorithm). Sometimes, dedicated hardware is used to speed up processing, for example, by employing graphics processing units (GPUs) for their massive parallel processing capability. The resultant extra cost and carbon footprint can become a consideration in decisions concerning the frequency of training, the choice of algorithm or the choice to employ machine learning altogether.

Models are deployed either in batch mode or in continuous mode, depending on whether the AI system has a direct need for the model results. Models in continuous mode typically have more strict performance efficiency requirements.

The second AI-specific particularity is that the organization should consider early in the life cycle the production data which the AI system will operate on. Considerations can include availability, fit for