Artificial Intelligence GUIDE

Artificial Intelligence GUIDE

Disclaimer:
This Guide is meant to assist life sciences companies in managing AI-enabled computerized systems in GxP regulated environments. This Guide is created and solely owned by ISPE. It is not a regulation, standard or regulatory guideline document. ISPE cannot ensure and does not warrant that a system managed in accordance with this Guide will be acceptable to regulatory authorities. Further, this Guide does not replace the need for hiring professional engineers, technicians, or consultants.

Limitation of Liability
In no event shall ISPE or any of its affiliates, or the officers, directors, employees, members, or agents of each of them, or the authors, be liable for any damages of any kind, including without limitation any special, incidental, indirect, or consequential damages, whether or not advised of the possibility of such damages, and on any theory of liability whatsoever, arising out of or in connection with the use of this information.

ISBN 978-1-946964-85-4

Preface

Technological advancements in Artificial Intelligence (AI) provide new opportunities for the life sciences. New forms of AI are demonstrating unprecedented capabilities, changing the system landscape across GxP regulated areas. While this technology is expected to continue to evolve and grow in complexity, many organizations are encountering challenges in adoption at scale.

This ISPE AI Guide provides a holistic framework for the effective and efficient use of AI. It supports organizations to achieve high-quality and compliant AI-enabled computerized systems. Serving as a bridge for the general concepts of *ISPE GAMP® 5: A Risk-based Approach to Compliance GxP Computerized Systems (Second Edition)* and the characteristics of AI, the Guide incorporates concepts previously developed in the industry as well as considers recent regulatory developments.

This is envisioned to be a valuable resource for GxP regulated companies, suppliers, service providers, regulatory agencies, and other interested parties involved in the design, development, implementation, and use of AI-enabled computerized systems. It intends to contribute to advancing industry practices by fostering AI literacy, facilitating effective collaboration, and promoting the use of critical thinking, thus enabling the adoption of new technology while helping to ensure patient safety, product quality, and data integrity.

Acknowledgements

The Guide was produced by an international team of volunteers, under the leadership and direction of:

Brandi M. Stockton (Lead)	The Triality Group, LLC	USA
Martin Heitmann (Co-Lead)	Independent Consultant	Germany
Eric Staib, MS, MBA (Co-Lead)	Syneos Health	USA

The work was supported by the ISPE GAMP Community of Practice (CoP) Software Automation and Artificial Intelligence (SA & AI) Special Interest Group (SIG).

Chapter Leads and Contributors

The Guide Leadership Team wishes to thank the following individuals for their contributions to the Guide; individuals marked with an asterisk (*) served as a lead for one or more chapters:

Rolf Blumenthal	Körber Pharma Software (Freelance)	Germany
Taylor Chartier*	Modicus Prime	USA
Aude Chetwynd	Technical and Professional Services LLC	USA
Joanne Donald*	Roche Products Ltd.	United Kingdom
Stephen Ferrell*	Valkit.ai	USA
Gail Francis	AstraZeneca	United Kingdom
Carsten Jasper*	Charles River Laboratories Inc.	Germany
Stuart Jones	MSD SBS Dublin	Ireland
Ian Lucas	Seer Pharma Pty. Ltd.	Australia
Michael Martone*	Apprentice.Io	USA
Stefan Münch*	Körber Pharma Consulting GmbH	Germany
Meher Muttanapalli*	Intuitive Surgical, Inc.	USA
Peyton Myers	Bürkert USA	USA
Tatum O'Kennedy	Seeq Corporation	USA
Raj Nandhan*	Medocity Inc.	India
Rick Rambo	Eli Lilly & Co.	USA
Mohammed Arif Rahman*	Otsuka Pharmaceutical Development & Commercialization Inc. (OPDC)	USA
Laila Rasmy, PhD*	University of Texas Health Science Center / FoundMed Co.	USA / Egypt
Doug Shaw	DShaw Consulting LLC	USA
Dr. Tomos Gwyn Williams*	Manchester Imaging Ltd.	United Kingdom
Kathy Zielinski	F. Hoffmann-La Roche Ltd.	Canada

Regulatory Input and Review

Thanks go to the members of regulatory agencies for their review and valuable comments on this Guide. The following members have agreed to be listed here:

Kevin Bailey	Medicines and Healthcare products Regulatory Agency (MHRA)	United Kingdom
Jason Wakelin-Smith	Medicines and Healthcare products Regulatory Agency (MHRA)	United Kingdom

Subject Matter Expert Input and Review

The AI Guide Leadership Team would also like to thank the ISPE Guidance Documents Committee (GDC) and GAMP Editorial Review Board (ERB) for their valuable comments on this Guide.

Special Thanks

The AI Guide Team recognizes Brandi M. Stockton (The Triality Group, LLC, USA) as the originator and driving force of the AI Guide initiative, whose leadership and vision laid the foundation for the development of this Guide.

The AI Guide Leadership Team would like to express particular thanks for Mark Newton (Heartland QA, USA), serving as a Mentor and providing insights from the ISPE GDC perspective, and Siôn Wyn (Conformity Limited, United Kingdom) for his guidance during the Guide's proposal stage.

The AI Guide Leadership Team would also like to thank the following staff members for their support during the development of this Guide:

- Nada Elsayed, PhD, ISPE Content Development Manager
- Lynda Goldbach, ISPE Sr. Publications Manager
- Gabriela Kantor, ISPE Technical Editor, Publications
- Rochelle May, ISPE Sr. Director, Publications
- Wendy McGee, ISPE Health Authority Outreach Manager
- Jeanne Perez, ISPE Guidance Document Technical Editor
- Tim Postlethwaite, PhD, Director, ISPE Technical Communities
- Nina Wang, ISPE Guidance Documents Technical Editor
- Heather Watson, ISPE Guidance Documents Technical Editor
- Carol Winfield, ISPE Sr. Director, Regulatory Operations

The ISPE Technical Editor Team wishes to thank Brandi M. Stockton and Martin Heitmann for their efforts and contributions during the editing process of this Guide; these individuals went above and beyond their responsibilities as leads in achieving clarity and consistency of the guidance.

The AI Guide Leadership Team would like to express their grateful thanks to the many individuals and companies from around the world who reviewed and provided comments during the preparation of this Guide; although they are too numerous to list here, their input is greatly appreciated.

Company affiliations are as of the final draft of the Guide.

Cover photo: www.istockphoto.com.

Table of Contents

Phase Appendices

Management Appendices

Special Interest Topics Appendices

General Appendices

3001 N. Rocky Point Dr. E., Suite 200-242, Tampa, Florida 33607 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

www.ISPE.org

1 Introduction

Advances in Artificial Intelligence (AI) offer new ways to approach challenges in GxP regulated processes. While such approaches offer promising advancements in support of various processes, further responsibilities arise in planning, developing, monitoring, and maintaining such systems that require thoughtful decision-making to ensure effective use of emerging technology.

While initial guidance is provided in the *ISPE GAMP® RDI Good Practice Guide: Data Integrity by Design* [1], and *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems (Second Edition)* [2], there is no single source of holistic guidance covering relevant concepts in the use of AI in GxP regulated areas in a comprehensive manner.

This Guide provides a holistic interpretation and operationalization of existing guidance and expectations regarding good practices in the context of AI-enabled computerized systems in GxP regulated environments and presents additional concepts for areas not covered. It aims to support alignment among diverse stakeholders to seize benefits from the quickly advancing technology by overcoming challenges in the adoption of AI-enabled computerized systems.

This Guide is part of the *ISPE GAMP® Guide Series* [3]. *"GAMP guidance aims to safeguard patient safety, product quality, and data integrity in the use of GxP computerized systems. It aims to achieve computerized systems, [including those that use AI,] that are fit for intended use and meet current regulatory requirements by building upon existing industry good practice in an efficient and effective manner. ...It does not represent prescriptive methods or a standard, but rather provides pragmatic guidance, approaches, and tools for the practitioner."* [2]

This Guide offers a robust, cost-effective approach when applied with expertise and good judgement. The approach is intended to be compatible with other frameworks, methods, and schemes, including:

- Quality systems standards and certification schemes, for example, the ISO 9001 series [4]
- ISO/IEC 42001: Information technology — Artificial intelligence — Management system [5]
- IEC 62304: Medical device software — Software life cycle processes [6]
- ISO 14971: Medical devices — Application of risk management to medical devices [7]
- ISO/IEC 23894: Information technology — Artificial intelligence — Guidance on risk management [8]
- Software process models such as ISO 12207: Systems and software engineering — Software life cycle processes [9]
- Schemes for assessing and improving organization capability and maturity, such as Capability Maturity Model Integration® (CMMI) [10]
- Iterative and incremental (Agile) software development methods and models
- Machine Learning (ML) life cycle processes, such as CRISP-ML(Q). The Life Cycle Process. [11]
- Approaches to IT service management, such as ITIL® [12]

Where possible, terminology is harmonized with standard international sources such as ICH [13] and ISO [14].

This Guide is intended as a stand-alone Guide. It is aligned with *ISPE GAMP 5 (Second Edition)*[1] [2] and is designed to be used in parallel with guidance provided in *ISPE GAMP 5 (Second Edition)* and other Guides in the *ISPE GAMP Guide Series* [3], such as *ISPE GAMP® Guide: Records and Data Integrity* [15]. This Guide aims to be compatible with the approach described in the ASTM E2500 [16].

1.1 Purpose

This Guide aims to safeguard patient safety, product quality, and data integrity in the use of AI-enabled GxP computerized systems, while also supporting the achievement of business benefit.

"Patient safety is affected by the integrity of critical records, data, and decisions, as well as those aspects affecting physical attributes of the product." [2] As in *ISPE GAMP 5 (Second Edition)*, the phrase "patient safety, product quality, and data integrity" is used throughout this ISPE AI Guide to underline this point.

This Guide provides a holistic view and good practices to achieve safe and effective use of AI-enabled computerized systems.

It provides a cost-effective good practice framework that helps to ensure that AI-enabled computerized systems are high-quality, effective, fit for their intended use, and compliant with relevant regulations. Key principles and practices relevant to the design, development, operation, and use of AI are presented.

This Guide is intended for use by regulated companies, suppliers, service providers, and regulators. Suppliers may be internal or external to the regulated company.

This Guide has been designed for use by a wide range of disciplines and responsibilities who share an interest in good practices for the use of AI that meets quality expectations and adheres to GxP regulations, including:

- Management
- Quality Unit
- Research
- Development
- Operations
- IT
- Support Staff
- Data Science
- Associated Suppliers and Service Providers

This Guide is not a standard or regulation. Regulated companies need to establish policies and procedures to meet applicable regulatory requirements.[2]

[1] *ISPE GAMP 5: A Risk-Based Approach to Compliant GxP Computerized Systems (Second Edition)* [2] is internationally recognized as the leading guidance document in this area.

[2] As stated in *ISPE GAMP 5 (Second Edition*) [2], *"it is inappropriate for regulated companies, suppliers, or products to claim that they are GAMP certified, approved, or compliant."*

1.2 Scope

This Guide applies to AI-enabled computerized systems covered by GxP regulations. GxP regulation are international requirements in the pharmaceutical and medical device ("life sciences") context. Requirements are imposed by international and national legislation or regulations under which a pharmaceutical or medical device company operates. Examples include:

- US Federal FD&C Act [17]
- US PHS Act [18]
- US FDA regulations [19]
- EU Directives [20]
- UK MHRA regulations [21]
- Japanese regulations [22]
- Prescription Drug Marketing Act (PDMA) [23]

Typical GxP areas those regulations refer to are [2]:

- Good Manufacturing Practice (GMP)
- Good Clinical Practice (GCP)
- Good Laboratory Practice (GLP)
- Good Distribution Practice (GDP)
- Good Quality Practice (GQP)
- Good Pharmacovigilance Practice (GVP)
- Medical Devices

Additional areas may apply.

This Guide is also consistent with other regulatory requirements such as:

- EU AI Act[3] [24]
- Data privacy regulations, e.g., EU GDPR [25] and US HIPAA [26]
- Sarbanes-Oxley (SOX)[4] [27]

This Guide provides an approach that is suitable for all types of computerized systems using AI, applicable to standard and configurable products as well as custom applications. Even though it focuses on AI-enabled GxP computerized systems, organizations may also take advantage of the guidance for other systems.

[3] The EU AI Act establishes a horizontal framework, including a risk-based AI classification system that imposes requirements with risk involved in the use of AI. It also calls out the need for AI literacy in Art. 4, which is discussed throughout this Guide.

[4] The US SOX Act (US), specifically Section 404, mandates control of computer systems that generate financial records. Many of the good practice principles and electronic records management controls are relevant to compliance with this law.

The principles in this Guide can be applied to a wide range of computerized systems using AI; however, not all activities described apply to all systems. Organizations should adopt a scalable approach (using critical thinking) to select the appropriate activities.

This Guide focuses on AI-specific components of a computerized system. It is not intended to repeat all the guidance from other GAMP Guides or Good Practices Guides. It contextualizes such guidance, and refers to others as applicable.

1.3 EU AI Act

The EU AI Act [24], adopted in 2024, is a horizontal regulation across industry sectors. It builds on the ethical principles developed in earlier guidance [28], such as trustworthy AI principles, to provide regulation on "AI systems." The EU AI Act defines dedicated roles ("operators") that are linked to key requirements:

- A "provider" in the sense of the EU AI Act is typically a "supplier" from the point of view of this ISPE AI Guide. On some occasions, a regulated company may also be considered as a provider, for instance for AI-enabled medical devices. Additionally, "importer" (importing an AI system from a third country into the EU market) or "distributor" (providing an AI system not as a provider nor as an importer) are typically specialized forms of suppliers in the sense of this ISPE Guide; exceptions may apply.

- The "deployer" defined in the EU AI Act is typically a regulated company in the context of this ISPE Guide.

At the core of the EU AI Act is a risk classification for AI systems comprising unacceptable risk (prohibited), high-risk (strict obligations), limited risk (transparency obligations), and minimal risk (voluntary code of conduct). Since the classification mechanism is highly specific, leveraging other regulations and classes of use cases, such assessment should be executed with care.

Even though terminology appears to be similar to established concepts in GxP areas, requirements should be carefully assessed when applying the EU AI Act. A system classified as limited risk per the EU AI Act may exhibit a high impact on patient safety, product quality, and data integrity.

Further regulatory requirements are imposed on so-called general-purpose models, including certain examples of Large Language Models (LLMs), depending on their characteristics.

The first linkage between GAMP concepts and the EU AI Act (and the proposal at the time of the article's publication [29]) was established in the *Pharmaceutical Engineering*® article "New EU AI Regulation and GAMP 5®" [30].

This Guide aims to provide good practices to help organizations achieve compliance with the EU AI Act; however, regulated companies need to consider specific requirements of the EU AI Act. Additional regional and local statutes apply.

1.4 Supplier Aspects

This Guide covers activities of the regulated company per the AI-enabled computerized system life cycle. This should not be confused with the defined approach or method for software development, which is the responsibility of the supplier.

This Guide defines activities and responsibilities expected of both internal and external suppliers of AI products and services. Internal suppliers should follow processes consistent with the regulated company Quality Management System (QMS).

Guidance on typical activities for the development of models is provided in this Guide, aiming for effective support of regulated company activities in their responsibility to determine the fitness for purpose of models regarding their context of use. However, the model development life cycle as described is not meant to be prescriptive, allowing suppliers to use the most appropriate methods and frameworks.

> *"Modern systems may have a complex supply chain involving multiple suppliers. This Guide aims to meet the needs of each group."* [2]

1.5 Business Benefits

Benefits of adopting approaches described in this Guide include:

- Focusing efforts based on the risk, thus allowing for efficient processes while achieving compliance
- Choice of system designs that fit the organization's experience
- Improved efficiency of collaboration among stakeholders, thus elevating the chances of successful AI implementation
- Effective monitoring and maintenance, demonstrating ongoing control of complex AI-enabled computerized systems
- Facilitating continual improvement of AI-enabled computerized systems during operation
- Facilitating efficient and effective collaboration of suppliers and regulated companies

1.6 Structure

This Guide forms part of a family of documents that together provide a powerful and comprehensive body of knowledge covering all aspects of computerized systems' good practices and compliance.

Separate Good Practice Guides cover the application of GAMP principles and frameworks to specific types of systems and platforms, while others detail approaches to specific activities and topics.

This ISPE AI Guide is comprised of a main body containing principles and a life cycle framework applicable to GxP regulated AI-enabled computerized systems. The main body is supported by several appendices providing practical guidance on a wide range of topics.

The key concepts described in Chapter 2 are the foundations on which this Guide is written. These concepts should be applied using critical thinking.

Chapter 3 introduces two life cycles, namely the AI-enabled computerized system life cycle and the model development life cycle including an AI sub-system. They include activities from the initial concept of understanding requirements, through purchase or development, release, and operational use, up to system retirement.

Chapter 4 describes activities occurring throughout life cycles in more detail, covering all life cycle phases: concept, project, operation, and retirement.

Chapter 5 provides a contextualization of Quality Risk Management (QRM) activities for AI-enabled computerized systems.

The organizational and governance framework, covered in Chapter 6, aids in ensuring effective and consistent regulated company activities to achieve compliance and fitness for purpose of AI-enabled computerized systems. It includes policies, responsibilities, management, and continual improvement.

The supplier has a key role to play in supporting compliance of the regulated company. Typical supplier activities are included in Chapter 7.

More detailed information is contained in a series of phases, management, and special interest topic appendices. Correlation of these topics with *ISPE GAMP 5 (Second Edition)* [2] is shown in Table 1.1.

Table 1.1: Alignment of ISPE AI Guide Appendices with ISPE GAMP 5 (Second Edition)

	AI Management (M)	AI Phases (P)	AI Special Interest Topics (S)
General Topics Extended to AI-Enabled Computerized Systems	• Quality by Design (QbD) • Supplier Management • Science-based QRM • Critical Thinking • Knowledge Management and Building AI Literacy • IT Infrastructure	• Concept Phase • Project Phase • Operation Phase • Retirement Phase	• Organizational Changes • Statutes, Regulations, Standards, and Guidance
Topics Specific to AI-Enabled Computerized System	• Fit for Purpose Data and Data Quality • Data and Model Governance and Management • Trustworthy AI • AI Maturity • Categories of Software and Hardware		• AI Adoption and Challenges • Machine Learning (ML) Fundamentals • Explainable AI (XAI) • Cybersecurity Challenges • AI in and as Medical Devices
General Computerized System Topics – refer to *ISPE GAMP 5 (Second Edition)*	• Validation Planning • Design Review and Traceability • Supplier Quality Planning • Validation Reporting • Project Configuration and Change Management • Documentation and Information Management • Alignment with Standards	• System Descriptions • Data Migration • Agile Software Development	

This Guide is intended to provide a flexible framework for achieving compliant AI-enabled computerized systems fit for intended use. Alignment with *ISPE GAMP 5 (Second Edition)* has been ensured, while focusing on aspects specific or highly relevant to the use of AI.

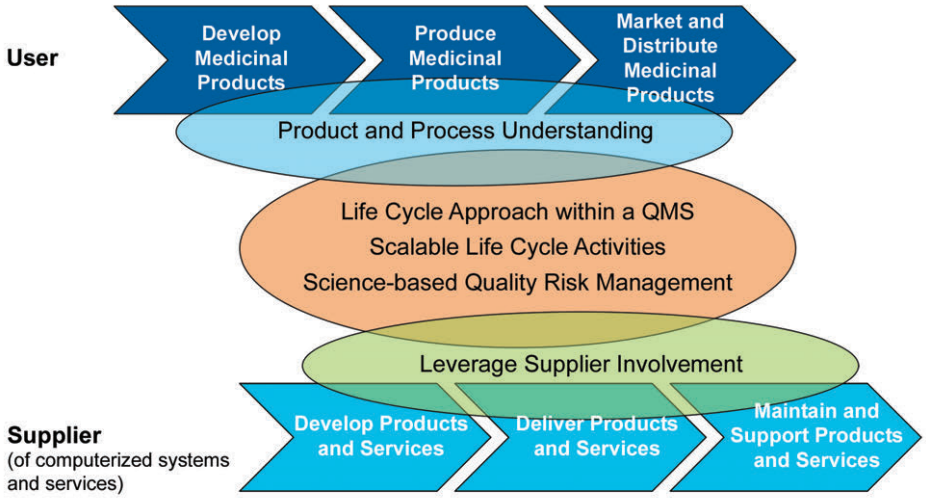
2 Key Concepts

2.1 Introduction

This Guide promotes the five key concepts of *ISPE GAMP 5 (Second Edition)* [2] as the foundation of AI-enabled computerized systems (shown in Figure 2.1):

1. *"Product and process understanding*
2. *Life cycle approach within a QMS*
3. *Scalable life cycle activities*
4. *Science-based QRM*
5. *Leveraging supplier Involvement"*

Figure 2.1: Key Concepts per *ISPE GAMP 5 (Second Edition)* [2]

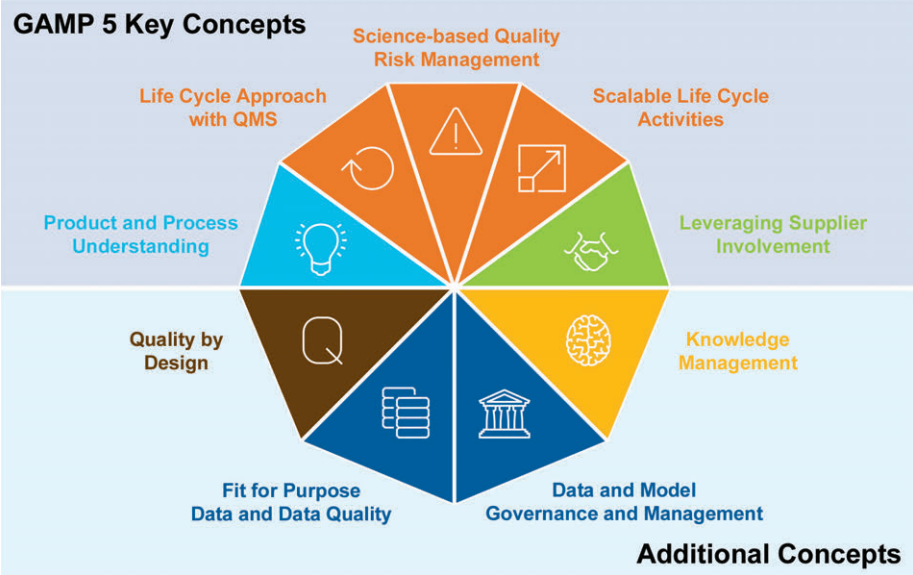


Four additional concepts are part of AI-enabled computerized systems:

- Quality by Design (QbD)
- Fit for Purpose Data and Data Quality
- Data and Model Governance and Management
- Knowledge Management

Together, these concepts constitute nine building blocks to achieve safe and effective AI-enabled computerized systems; see Figure 2.2.

Figure 2.2: ISPE AI Guide Concepts Overview



2.2 *ISPE GAMP 5 (Second Edition)* Key Concepts Overview

A brief overview of *ISPE GAMP 5 (Second Edition)* [2] key concepts is provided, including a contextualization for AI-enabled computerized systems.

2.2.1 Product and Process Understanding

Product and process understanding serve as the basis upon which computerized systems are built that are fit for purpose and deemed safe for operation.

ICH Q10 [35] renders product and process knowledge foundational to the pharmaceutical quality system: *"Product and process knowledge should be managed from development through the commercial life of the product up to and including product discontinuation." ISPE GAMP 5 (Second Edition)* [2] provides a wider view to product and process understanding in its fundamental role to determine system requirements and achieve data integrity.

Such understanding enables:

- Clear definition of intended use and requirements
- Performing risk management and assessment(s), leveraging critical thinking
- Design of an AI-enabled computerized system that fulfills intended use
- Articulation of the computerized system's purpose to all relevant stakeholders according to their needs
- Maintaining a state of control

2.2.2 Life Cycle Approach Within a Quality Management System

A life cycle approach structures and integrates activities from concept to retirement of the AI-enabled computerized system with its supporting processes. It aims to achieve and maintain the safety and efficacy of AI-enabled computerized systems throughout the life cycle and aid in coordinating stakeholders and their individual contributions.

It relies on the concepts of specification and verification to ensure quality of the AI-enabled computerized system, while considering development aspects specific to AI, such as iterative experimentation.

Further information on a suitable life cycle approach for AI-enabled computerized systems is provided in Chapter 3, with additional guidance in Chapter 4 and Appendices P1, P2, P3, and P4.

2.2.3 Scalable Life Cycle Activities

When managing an AI-enabled computerized system, decisions on the scale of activities are needed; see also ICH Q9(R1) [31]. Life cycle activities should be scaled according to system impact, system complexity and novelty, and outcomes of supplier assessment. The business impact may also influence the appropriate level of activities [2].

Examples of scalable life cycle activities specific to AI-enabled computerized systems include:

- Rigor applied to records and information
- Amount of effort invested in performing data and model engineering activities
- Testing efforts of models and AI-enabled computerized systems
- Effort invested in user training prior to and during the operation phase

Similar considerations are needed for supporting processes, such as determining the right level of effort for activities. Apply critical thinking to achieve this outcome.

Appendix M4 provides guidance on the application of scalable life cycle activities in seven areas: Concept, Project, Operation, Retirement, Risk Management, Supplier Assessment and Audit, and Inspection Readiness.

2.2.4 Science-Based Quality Risk Management

"QRM is a systematic approach for the identification, assessment, control, communication, mitigation, and review of risks." [2] It allows for decision-making based on risk in a justified manner using critical thinking.

The following key risk management principles support efficient achievement of safe and effective AI-enabled computerized systems:

- Science-based approach founded in empirical evidence
- Objective decision-making based on available knowledge[5]
- Forward-looking approach starting early in a project
- Leveraging experience and knowledge
- Managing diverse expertise to arrive at a holistic risk perspective

Chapter 5 and Appendix M3 provide guidance on managing risks throughout the AI-enabled computerized system life cycle.

2.2.5 Leveraging Supplier Involvement

Often, regulated companies involve suppliers in their activities to develop, use, and maintain computerized systems. Effective collaboration is a key consideration to achieve efficiency and high-quality outcomes.

[5] One driver to update ICH Q9 [31] was the need to strengthen the relationship between risk management and knowledge management.

When leveraging supplier involvement, regulated companies should be aware of their accountability, as stated in *ISPE GAMP 5 (Second Edition)* [2]: *"Responsibility for activities may be with the suppliers, but in all cases regulatory accountability lies with the regulated company."* Regardless of who performs an activity, there should be no decrease in product quality, process control, or quality assurance.

Further guidance is in Chapters 6 and 7 and Appendix M2.

2.3 Additional Concepts Overview

2.3.1 Quality by Design (QbD)

A robust quality system is foundational to establishing computerized systems fit for intended use and for demonstrating ongoing control. It includes:

- Systematic evaluation and understanding of the process and goals in the context of use
- Establishing a scientific based, forward-looking QRM approach
- Proposal for design spaces that fit the process and its goals
- Use of experimentation to identify the relationship between inputs and process data on performance indicators
- Thorough testing to provide evidence of fitness for purpose prior to operation

This Guide adopts a forward-looking approach to quality, leveraging QbD principles and specific quality dimensions relevant to AI-enabled systems, and embeds life sciences specific aspects of the quality system in line with ICH guidance[6]; see Appendix M1.

2.3.2 Fit for Purpose Data and Data Quality

Data plays a significant role in achieving safe and effective AI-enabled computerized systems. Regulated companies require data that is fit for purpose in the context of use, which means that considered data meets four properties: reliability, relevance, representativeness, and abundance. Data understanding allows for deriving possibilities and limitations of the use of data in such systems.

Further guidance on fit for purpose data and data quality is provided in Appendix M6, and in life cycle guidance throughout the Guide.

2.3.3 Data and Model Governance and Management

Data and model governance and management are the organizational structures, policies, standards, and procedures concerned with data and models as the baseline for safe and effective AI-enabled computerized systems.

Robust and documented data and model governance and management practices help to achieve successful implementation and safe operation of any computerized system.

Regulated companies should consider data and model governance at an organizational level rather than isolated on a single use case. This supports scaling use cases across the organization, reusability of data and models, and leveraging the effort invested while reducing risks.

Further guidance is included in Appendix M7.

[6] For instance, ICH Q9(R1) [31], ICH Q10 [35], ICH E6 [36] and ICH E8 [37] may be considered.

2.3.4 Knowledge Management

The *ISPE Good Practice Guide: Knowledge Management in the Pharmaceutical Industry* [38] definition of knowledge management[7] is:

> *"The application of a structured process to help information and knowledge flow to the right people at the right time so they can act more efficiently and effectively to find, understand, share, and use knowledge to create value."*

Regulated companies should consider specific aspects in knowledge management to achieve and maintain effective and high-quality AI-enabled computerized systems. Specifically, AI literacy is a key facilitator for the efficient implementation, effective adoption, and ethical use of AI-enabled computerized systems, relying on the combination of domain with data science expertise. Further guidance is provided in Appendix M5.

2.4 Key Terms

Key terms relevant to AI-enabled computerized systems are described in Sections 2.4.1–2.4.7. Further terms from *ISPE GAMP 5 (Second Edition)* [2] are applicable:

- **Computerized System:** *"A computerized system consists of the hardware and software components, together with the controlled function or process (including procedures, people, and equipment and associated documentation)."* See Section 2.4.2 for the extension to AI-enabled computerized systems.
- **Computerized System Validation:** *"Achieving and maintaining compliance with applicable GxP regulations and fitness for intended use by the adoption of principles, approaches, and life cycle activities within the framework of validation plans and reports [and] the application of appropriate operational controls throughout the life of the system."* See Section 3.4.
- **GxP Regulation:** *"The underlying international pharmaceutical requirements…or other applicable national legislation or regulations under which a company operates."* See Section 1.2.
- **GxP Compliance:** *"Meeting all applicable pharmaceutical and associated life-science regulatory requirements."*
- **GxP Regulated Computerized System:** *"Computerized systems that are subject to GxP regulations."*
- **Process Owner:** *"The owner of the business process or processes being managed, [who] is ultimately responsible for ensuring that the computerized system and its operation is in compliance and fit for intended use in accordance with applicable company Standard Operating Procedures."* See Chapter 6 and Appendix G2.
- **QMS:** *"Management system to direct and control an organization with regard to quality."* See relevant ISO standards, such as ISO 9001 Quality management systems — Requirements [4].
- **Subject Matter Expert (SME):** Individuals with specific expertise in a particular area or field.
- **System Owner:** *"The system owner is responsible for the availability and support and maintenance of a system, and for the security of the data residing on that system."* See Chapter 6 and Appendix G2.

2.4.1 Data Science, Artificial Intelligence, and Machine Learning

The hierarchical relationship between data science, AI, and ML is described in the following subsections.

[7] Another valuable resource on knowledge management is the American Productivity & Quality Center (APQC) [39].

2.4.1.1 Data Science

Data science combines math and statistics, specialized programming, advanced analytics, and AI with subject matter expertise to uncover actionable insights unseen in an organization's data. Data science provides techniques for preparing and analyzing the data used to develop AI systems and train models based on ML [40].

2.4.1.2 Artificial Intelligence

Understanding AI becomes easier when seen through the lens of human intelligence. Psychologists generally characterize human intelligence as a combination of abilities, including reasoning, discovering meaning, generalizing, or learning from experience. Mirroring some of these abilities, AI often focuses on these main components: learning, reasoning, problem solving, perception, and using language. [41]

Regarding AI used in a system, the OECD [42] states: *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."* This is also the definition presented by EMA in their reflection paper [43].

The FDA [44] has proposed: *"A branch of computer science, statistics, and engineering that uses algorithms or models that exhibit behaviors such as learning, making decisions, and making predictions."*

Similarly, the European Parliament [45] suggests: *"AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity."*

As presented in the US Executive Order[8] [46], AI is defined as *"a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action."*

The European Commission interprets the terminology of an AI system in the context of the EU AI Act [24, 47], with exclusions that, according to these guidelines, would not be considered pertaining to the requirements of the EU AI Act. This ISPE Guide considers a wider interpretation of models that would constitute the use of guidance to achieve safe and effective AI-enabled computerized systems helpful.

AI may also be understood according to the tasks to which it has been successfully applied. Examples of AI capabilities are:

- **Extracting information from unstructured data:** LLMs allow extraction of data and metadata from previously unstructured data (e.g., text or images) that is not accessible via traditional or rule-based approaches. For example, it allows for categorization or summary of texts, findings, and scientific publications, such as automating deviation categorization.

- **Recognizing complex patterns:** The ability to identify patterns by learning from data was introduced in *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* Appendix S1 [1], with information added in *ISPE GAMP 5 (Second Edition)* [2] regarding clustering and forecasting. It also enables the identification of patterns that are not accessible using traditional statistical methods. In conjunction with a model to extract information from unstructured data, example purposes include the identification of potential side effects in pharmacovigilance for the improvement of patient safety or assisting in the identification and application of clinical guidelines.

[8] The order was revoked in 2025 and is listed here for historical purposes, providing an overview of various approaches to define AI and AI systems.

- **Real-time automated image processing:** A class of neural networks known as Convolutional Neural Networks (CNNs) are specifically designed to process and generate images; see also Appendix S3. The effectiveness of CNNs has improved significantly in recent years due to the availability of increasingly powerful computation hardware. Starting with simple networks only able to process uncomplicated images e.g., hand-written post codes, CNNs have evolved to include 50 or more layers processing complex images to identify complex structures. Current CNNs allow for real-time processing of images much faster and more reliably than human eyes. An advantage of using this method in fill and finish processes is for the quality control of filled vials as it significantly enhances the safety of the final drug product.

- **Managing complex knowledge:** AI, in particular semantic analysis, supports scientists, engineers, and regulatory specialists in handling problems that require a high level of knowledge, e.g., Root Cause Analysis (RCA) in deviations.

- **Decision support:** Decision support systems and expert systems have existed for decades. These systems are primarily based on classic simple decision trees and other methods with varying complexity allowing the use of classic probability calculations. The integration of AI methods allows for significantly more complex decision support systems that permit more determination of probabilities of binary decisions or for various scenarios. Combined with CNNs for image evaluation, they can, for example, determine a risk score (probabilities) for certain diseases based on images from imaging procedures.

- **Generating Human-like content from multiple sources:** Generative AI has provided new ways to elicit human-like output via application of complex models. These can be used for a variety of tasks, from a brainstorming assistant to more formal use cases, such as translation of product information and labels, information extraction for case intake processes in pharmacovigilance, and effectively browsing complex sources of information (such as a set of regulatory requirements or SOPs).

2.4.1.3 Machine Learning

The definition of ML used in this Guide is from ISO [48]: *"Machine learning (ML) is a type of artificial intelligence that allows machines to learn from data without being explicitly programmed."*

ML covers all types of models that learn from existing data to infer real-world relationships between model input and model output, encapsulating such relationships in a (complex) mapping. This uses learning strategies, enabling them to establish this mapping and derive predictions or decisions based on new data.

Like the concept of AI, many definitions have been established to describe ML. For example, *"Machine learning (ML) is a branch of artificial intelligence (AI) focused on enabling computers and machines to imitate the way that humans learn, to perform tasks autonomously, and to improve their performance and accuracy through experience and exposure to more data."* [49]

The FDA defines ML in the discussion paper Artificial Intelligence in Drug Manufacturing [44]: *"A branch of AI that provides systems with the ability to develop models through analysis of data without being explicitly programmed and to improve based on data or experience."*

EMA states in their reflection paper [43] that *"Machine learning refers to the computational process of optimising the parameters of a model from data, which is a mathematical construct generating an output based on input data. Machine learning approaches include, for instance, supervised, unsupervised and reinforcement learning, using a variety of methods including deep learning with neural networks."*

In the US Executive Order [46][9], ML is defined as *"A set of techniques that can be used to train AI algorithms to improve performance at a task based on data."*

Common to all approaches is the idea that data informs a model that captures relations observed therein.

[9] The order was revoked in 2025 and is listed here for historical purposes, providing an overview of various approaches to define AI and AI systems.

ML can use a relatively simple construction such as regression, or more complex ones such as a neural network that mimics the functioning of a human brain. However, such models share similar characteristics as demonstrated by the example of linear regression analysis and a neural network; see also Appendix S3:

- **Regression analysis** uses several X/Y pairs (coordinates) to define a function that can map the position of all X/Y pairs as accurately as possible. This function can then be used to determine a Y for any given X. The examples lead to a general function that can then be used for other values.

- The same approach is applied when using a **neural network** trained with historical data to infer predictions regarding a particular objective. Given the complexity of model architecture, more complex, even multimodal data, can be handled (e.g., a combination of images and natural language).

While many forms of AI rely on ML, forms of non-ML AI exist. For instance, some forms of Natural Language Processing (NLP), such as a similarity-based knowledge retrieval system, may not necessarily rely on ML approaches, although they could be considered as AI.

2.4.2 Foundation Models

Foundation models have gained particular interest in recent years. Foundation models are large, deep learning neural networks trained on massive data sets. Rather than developing models from scratch, a foundation model can be used as a starting point for use in AI sub-systems and may provide cost advantages.

The term foundation model was coined by researchers to describe models trained on a broad spectrum of generalized and unlabeled data, capable of performing a wide variety of general tasks such as understanding language, generating text and images (hence called Generative AI), and conversing in natural language. BERT, GPT, CLAUDE, TITAN, Jurrasic-1, BLOOM are examples of foundation models.

While foundation models provide a means to leverage learning from a large data set and thus multiply the impact and benefits from the use of extensive resources for the initial training of the model across various stakeholders, several challenges apply:

- **Integration:** For practical use in AI-enabled computerized systems, developers need to integrate foundation models into a software stack, including possible elements such as prompt engineering, fine-tuning, and pipeline engineering.

- **Comprehensibility:** Foundation models that generate text can provide grammatically and partially correct answers. However, it may be difficult to provide the full context in a prompt, while a perfect response to the context cannot be expected either. For example, foundation models are generally not viewed as socially and psychologically competent.

- **Reliability:** Answers to questions may be unreliable and sometimes inappropriate or incorrect.

- **Bias:** Bias may arise as models might embody inappropriate patterns from training data sets.

- **Complexity of fine-tuning:** Many foundation models offer the possibility to fine-tune the model with an organization's own data to improve accuracy for a specific task. However, this approach comes at the cost of higher model maintenance complexity, given the interplay of the foundation and the fine-tuned model and their iterations derived from new data.

Foundation models can provide powerful capabilities regarding processing and generation of natural language, in the form of LLMs. When integrating LLM, Retrieval Augmented Generation (RAG) is a common technique with the following features (see Figure 2.3):

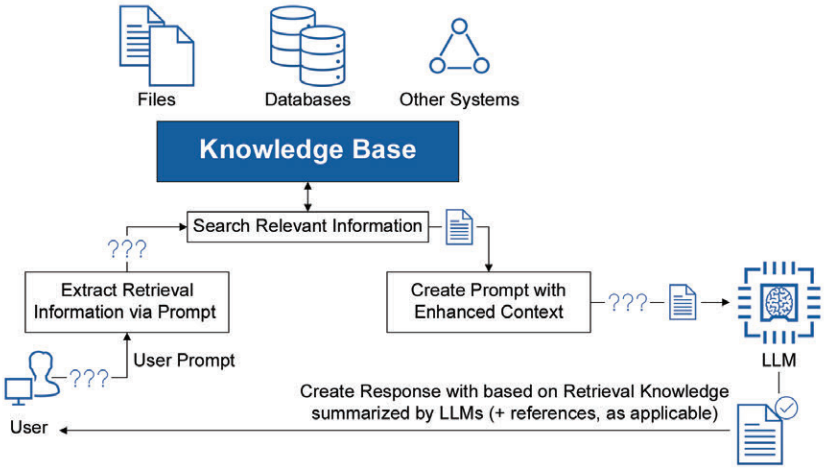
- Knowledge retrieval: Given a question of interest, the system extracts information from a predefined knowledge base to derive valuable case specific model input for the LLM.

- Prompting: The prompt integrates the original question of interest with information retrieved from the knowledge base.
- Generation of the answer: The LLM generates an answer based on the prompt.
- Post-processing: Model output may be processed further.

This approach provides various advantages:

- New information can enrich the knowledge base, so that the AI sub-system may provide answers based on up-to-date information without the need to train or fine-tune the models.
- Specific content can be integrated to inform and guide the LLM.
- References can be extracted, based on the retrieved information from the knowledge base to support the user of the AI-enabled computerized system in verifying the model output.

Figure 2.3: Schematic Representation of Retrieval Augmented Generation (RAG)



Foundation models, LLMs, and RAG approaches can be used for even more complex system designs, such as agentic AI. Agentic AI employs one or more agents with dedicated roles, tasks, and tools that can be used to support complex, abstract processes such as knowledge gathering, research, or analysis of data in a flexible manner.

2.4.3 AI-Enabled Computerized Systems

PIC/S guidance [50] provides a general overview of **computerized system's components** with a contextualization in the use of AI:

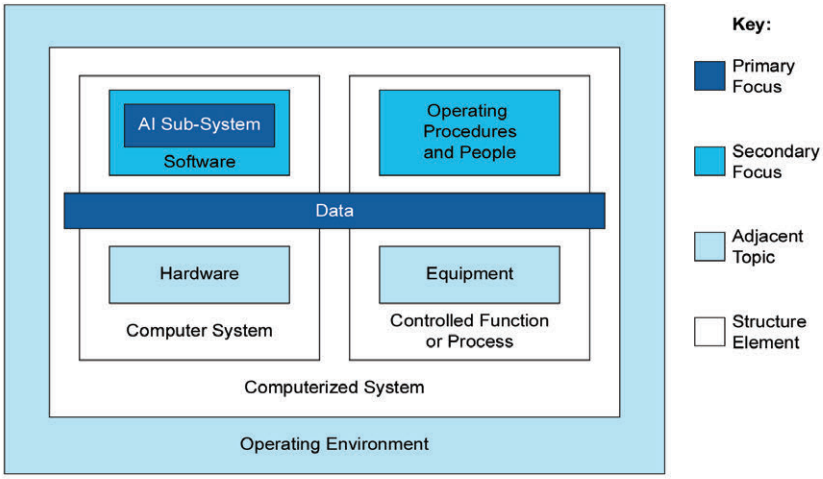
- **Computer System:**
  - **Software:** Software development and quality assurance considerations are the primary focus of this Guide, establishing the concept of an AI sub-system as a module of the AI-enabled computerized system that includes AI; see also Section 2.4.3.
  - **Hardware:** Considerations on hardware are discussed, given the potentially high hardware requirements for model engineering, model use, and auxiliary functionality to support human-AI interaction, such as explainable AI (XAI) methods.

- **Controlled Function or Process:**
  - **Operating Procedures and People:** The relevance and necessity of human oversight and control of AI-enabled computerized systems regarding processes and activities throughout the life cycle.
  - **Equipment:** Equipment is covered to the extent that the AI-enabled computerized system and its AI sub-systems influences or controls equipment.
- **Operating Environment:** While not a focus of this Guide, various aspects of the operating environment, such as knowledge and organizational management, are covered to establish best practices.
- **Data:** Data is an essential component of an AI-enabled computerized system, connected to all stages of the life cycle. Data intersects with all other areas mentioned here, including data storage "at rest" in hardware, during processing in software, in interaction with human operators, and as aggregated to facilitate risk-based decision-making rooted in product and process understanding as well as data understanding.

To distinguish AI-enabled computerized systems from those that do not make use of AI, the term non-AI-enabled computerized system is used.

Figure 2.4 summarizes the focus areas of this Guide.

Figure 2.4: Computerized System Overview and Focus Areas of this Guide



2.4.4 Model, AI Sub-System, AI Function, AI System, and AI-Enabled Computerized System

A model is a sub-program that processes model inputs and derives model outputs. Models aim to capture the essence of the real-world relationship between such model inputs and the ground truth, i.e., the desired model output consistent with the real-world to serve a dedicated task.

Models may include ML ("ML models") or can be based on a complex set of rules ("non-ML models"), and can serve a variety of tasks, such as clustering, classification or regression, and generation of text or other data; see Appendix S3.

For productive use, models need to be integrated into a computerized system, then into an AI-enabled computerized system that supports their operation. To integrate a model, the following layers are typical (see also *ISPE GAMP 5 (Second Edition)* Section 23.3.3.5 [2]):

- The model including pre-processing of model input and post-processing of model output (if applicable) is a module called **AI sub-system.**

- AI sub-systems are part of Functions, which are then called **AI Functions.** They may contain one or more AI sub-systems and other non-AI modules, such as data sourcing, presentation of model output, and allowing for interaction with end users.

- AI Functions are part of Components, or computer systems, called an **AI System.** An AI system may include one or many AI Functions and non-AI Functions.

- AI Systems are integrated into a computerized system, called an **AI-enabled computerized system.** An AI-enabled computerized system may hold one or many AI Systems, non-AI computer systems, and additional components, as shown in Figure 2.4.

2.4.5 Iterative Experimentation

Iterative experimentation is a process to derive models, including pre- and post-processing functionality, aiming to achieve the best-performing model given one or a set of Key Performance Indicators (KPIs). An experiment consists of the following steps:

1. Model selection – selecting a model including its model input, hyperparameters, and configurations

2. Data engineering – implementing pre- and post-processing functionality

3. Model engineering – establishing the model (e.g., configuration, coding, training, fine-tuning)

4. Model evaluation – measuring model performance to guide next development steps

The evaluation may either lead to continuation of iterative experimentation or closure of the iterative experimentation process.

The iterative nature of the model development life cycle should not be confused with the incremental building of software. While model development aims for improvements of models that serve one specified objective, incremental building of software primarily aims for stepwise implementation and verification of subsequent software functions. However, both use “agile” approaches. (See Appendix D8 of *ISPE GAMP 5 (Second Edition)* [2] and the Manifesto for Agile Software Development [51].)

2.4.6 Relation Between Key Life Sciences and Data Science Terms

Life sciences, covering GxP areas, and data science, a horizontal concept, have developed independently. Therefore, specific terminology has been established within each domain, including some similarities in choice of terminology.

In the life sciences, dedicated terms have been established to describe activities and concepts to ensure a computerized system’s fitness for intended use:

- **Validation** includes all activities performed to determine a computerized system’s fitness for intended use and for achieving compliance with applicable regulations. Through validation, documented evidence is established by which a high degree of assurance is provided that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes [52].

- **Verification** is the *“confirmation, through the provision of objective evidence, that specified requirements have been fulfilled.”* [2]

- **Testing** describes a selection of strategies that apply to a system’s capabilities, or those of a lower sub-system or component of the system, to verify that its behavior is in line with the design, or its specification depending on the testing level. According to *ISPE GAMP 5 (Second Edition)* Section 7.10 with further guidance provided in Appendix D5, testing may include module (unit) testing, integration testing and system testing.

Data Science developed concepts to derive meaningful indications of a model's performance, by its application to suitable data. Terminology evolved around splitting data for various purposes:

- The **training data set** is used to derive an ML model, if applicable. It is not linked to the above life sciences terminology, though it is listed here for completeness.

- The **validation data set** is used to evaluate the performance of a model, based on unseen data. It is an inherent part of the incremental development process and is typically used many times. While it also provides insights into characteristics such as performance loss in comparison to those derived from training data, it is not seen as a means of formal verification.

- The **test data set** is used after iterative model development, independent from the training set (if applicable) and the validation data set. Applying the model to the test data set to derive KPIs can be seen as a verification activity, as the model requirements specifications are verified by this use of independent, unseen data; this is called model testing in this Guide.

Given this overview, the crossing of "validation" may cause confusion between data science and life sciences terminology. More specifically, the term "validation set" may carry a particular meaning, for example in regulatory requirements for validation data in medical devices [53, 54]; hence the validation data set (in a data science sense) is called tuning data set at times. Similarly, the FDA [55] makes use of the tuning data terminology, while also calling the combined training and tuning data "development data."

However, this Guide uses the current industry practice of training, validation, and test data sets. When it is not clear from the context, a qualifier is added to the validation data set to avoid confusion.

Further information on those data sets and data splitting in general is found in Appendix P2.

2.4.7 Static and Dynamic Systems

Dynamic and static systems refer to the adaptiveness of models in operation and are characterized as follows:

- Dynamic systems are designed by which embedded models may evolve automatically without approval by a human operator.

- Static systems remain on a dedicated version without change of parameters unless a manually executed change of version is performed.

Static models are more commonly implemented at the time of this publication, primarily because of simpler data and infrastructure requirements, as well as validation efforts to maintain a state of control.

However, dynamic system designs may help the model to stay on track in case of changing environments and dynamic relationships in or between input and output. For illustrative purposes, dynamic systems may be considered when:

- Additional case data is expected to improve the model's performance during runtime, to expand compatibility and increase generalization for further application within the system

- Shifts in human interaction with the model or added information that needs to be reflected dynamically (e.g., change in treatment procedures) are expected in operation

- Trending is expected, which may result in loss of performance because the training and test set is no longer representative ("model drift," "bias," and challenges to fairness)

Various forms of dynamic AI-enabled systems exist. Some can be seen as automatized development, testing, and deployment (e.g., retraining a regression model based on new data; controlled by time, by amount of new data, or by KPIs). In other cases, dynamic learning is an integral part of the model type (e.g., forms of reinforcement learning). Further guidance on various degrees of autonomy and adaptiveness is provided in Appendix M10.

Choosing a dynamic operating model has substantial implications for the system's life cycle activities, as outlined throughout this Guide. Particularly, risks that not only pertain to the model itself, but also its incremental learning path need to be considered; see Appendices P1, P2, and P3.

Of note, regulated companies should consider (local) regulatory requirements and guidance; for example, EMA excludes the use of dynamic systems in pivotal clinical trials [43].

2.4.8 Trustworthy AI

Trustworthy AI considers AI and its use with human values, related to responsible AI and ethical use of AI. Various frameworks exist that address trustworthy AI. While these frameworks share common concepts, nuances in local guidance should be considered.

For instance, the European Commission [28] defines trustworthy AI as AI that is:

- Lawful, complying with applicable laws and regulations such as the EU AI Act (see Statutes, Regulations, Guidance and Regulatory Initiatives)
- Ethical, ensuring adherence to ethical principles a) respect for human autonomy, b) prevention of harm, c) fairness, and d) explicability
- Robust, both from a technical and social perspective

The European Commission underpins their concept of trustworthy AI by seven requirements [28]:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountability

As an additional example, the Australian government has defined AI ethics principles designed to ensure safe, secure, and reliable AI [56]:

- Human, societal, and environmental wellbeing
- Human-centered values (human rights, diversity, and the autonomy of individuals)
- Fairness
- Privacy protection and security

- Reliability and safety
- Transparency and explainability
- Contestability
- Accountability

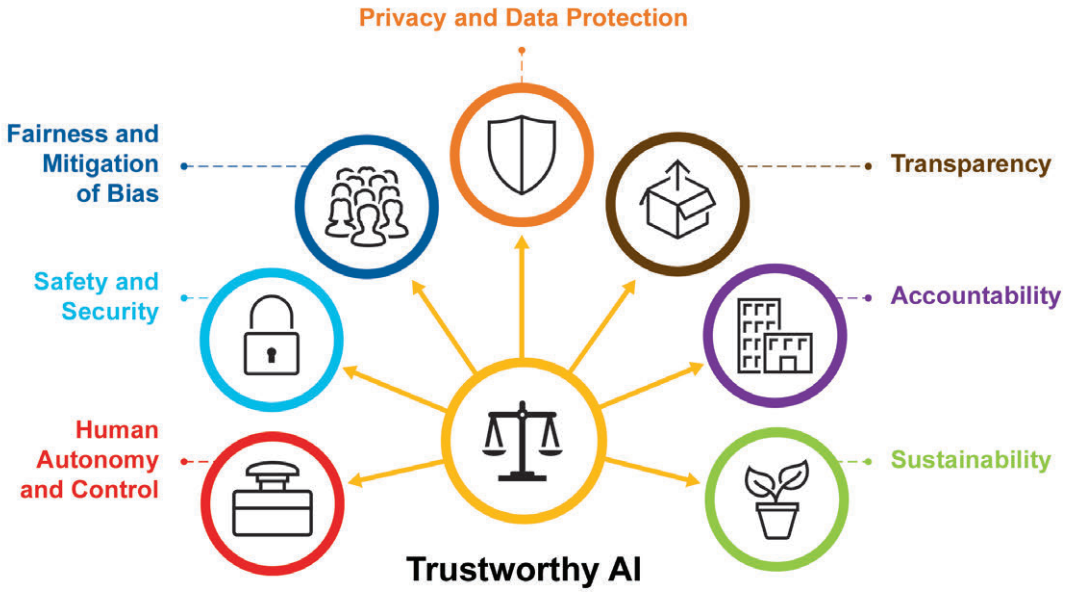
Tensions can arise between trustworthy AI considerations. For example, when defining the target populations for an AI-enabled medical device, excluding patient populations may be seen unfair, as the excluded populations will not receive treatment. But it may be necessary if there is no evidence that the device is safe and effective for the excluded population.

This ISPE Guide applies the following principles of trustworthy AI for consideration in AI-enabled systems (see Figure 2.5):

- **Human autonomy and control:** Assurance of humans' ability to effectively interact with and supervise AI
- **Safety and security:** Minimization of potential harm to users, patients, products, and equipment
- **Fairness and mitigation of bias:** Reducing algorithmic bias to improve quality of the AI model output across segments in the context of use
- **Privacy and data protection:** Protection and respect for personal data
- **Transparency:** Comprehensive decision-making and effective communication
- **Accountability:** Ensuring that organizations are held accountable for patient safety
- **Sustainability:** Consideration of ecological, societal, and governance implications

See Appendix M9.

Figure 2.5: Principles of Trustworthy AI



3 Life Cycle Approach

3.1 Introduction

Fitness for intended use and regulatory compliance may be achieved by adopting a life cycle approach [2]. A life cycle approach is a systematic way of modularizing and following steps in a logical order to facilitate achieving high-quality computerized systems.

Good engineering practices as described in ASTM E2500 [16] underpin the life cycle approach; see also *ISPE GAMP 5 (Second Edition)* [2]. It includes the concept of specification, design, and verification to ensure that computerized systems and their components meet requirements, leading to acceptance and release, while maintaining a state of control during operation.

3.2 Life Cycles

Two life cycles are the focus of this Guide:

- An **AI-enabled computerized system life cycle,** supporting the computerized system from conceptualization to retirement
- A **model development life cycle** including an AI sub-system, which integrates the perspective of the AI-enabled computerized system with model development activities

This Guide describes the overall AI-enabled GxP computerized system life cycle from the perspective of the regulated company; the model development life cycle is described from the perspective of a supplier (internal or external to the organization). The AI sub-system represents an interface to connect the model development life cycle with the AI-enabled computerized system life cycle and addresses the regulated company and the supplier.

These life cycles are not a substitute for the software development approach of the supplier. Further information on the main elements of the three life cycles is provided in the following sections. Chapter 4 provides details on their integration and their activities.

3.2.1 AI-Enabled Computerized System Life Cycle

The AI-enabled computerized system life cycle follows general considerations provided in *ISPE GAMP 5 (Second Edition)* [2], including four phases:

- **Concept Phase:** *"During the concept phase, the regulated company should consider opportunities to automate one or more business processes based upon business need and benefits."* [2] Upon identification of a business need or opportunity, activities such as engaging stakeholders and an initial risk assessment are performed. Having identified potential solution approaches, the use of AI may be considered, which leads to the definition of the context of use and the scope that should be served by the AI sub-system. Further activities related to the AI sub-system apply; see Section 3.2.2. Establishing initial requirements prepares the transition to the project phase.
- **Project Phase:** The project phase utilizes results from the concept phase and aims for a successful release of the AI-enabled computerized system. It includes *"planning, supplier assessment and selection, various levels of specification, configuration (or coding for custom applications), and verification leading to acceptance and release for operation"* [2], based on a validation report. Risk management is applied throughout the project phase to achieve an acceptable level of risk. Specific to the project phase of AI-enabled computerized systems is the integration of one or many AI sub-systems; see also Section 3.2.2 and Chapter 4.

- **Operation Phase:** The purpose of this phase is to achieve the benefits anticipated using AI while maintaining a state of control and fitness for intended use during operation. The operation phase uses *"defined, up-to-date, operational processes and procedures applied by personnel who have appropriate training, education, and experience."* [2] While incremental changes may apply, in the same way as non-AI-enabled computerized systems, characteristics for AI-enabled computerized systems are the collection of data during operation, which may lead to changes to AI sub-systems per change management procedures. The decision to retire the AI-enabled computerized system concludes this phase.

- **Retirement Phase:** The termination of the AI-enabled computerized system and their AI sub-systems occurs in this phase. This phase involves *"decisions about data retention, migration, or destruction, and the management of these processes."* [2] It includes planning and reporting on retirement activities.

Suppliers should be involved throughout the life cycle, as appropriate.

An inventory of computerized systems should be maintained; see Chapter 6.

An initial risk assessment should be conducted prior to the project phase. This includes assessment of regulatory impact, determination of applicable regulations, and identification of specific components to which these regulations apply to inform the planning in the project phase. See Step 1 of Chapter 5.

Seeking continual improvement accompanies the life cycle, allowing for reflection on insights gained to translate to enhanced product and process understanding, as well as data understanding.

Chapter 4 provides further details on typical activities performed throughout the AI-enabled computerized system life cycle, while Appendices P1, P2, P3, and P4 offer more detailed guidance.

3.2.2 Life Cycle Model Including an AI Sub-System

The life cycle model (shown in Figure 3.1), including an AI sub-system, is organized in parallel to the phases of the AI-enabled computerized system life cycle:

- **Concept Phase:** Provided that the use of AI is foreseen, a Proof of Concept (PoC) is planned, which includes a feasibility assessment to address the availability and suitability of data, and the identification of potential models. The concept phase may also include the development of a prototype as a simplified version of the AI sub-system. Initial insights into the interplay of data and models inform requirements and the decision whether to continue development activities in the project phase.

- **Project Phase:** Model requirements specifications are derived that inform subsequent activities for the selection of data and model development. They allow for verification of the model by means of model testing. Results from model testing inform additional testing activities and the decision for acceptance of the AI-enabled computerized system.

- **Operation Phase:** Ongoing performance monitoring leverages new data gathered during operation of the system, which generates performance indicators. Performance indicators and new data may lead to iterations of the AI sub-system, seeking to improve performance or mitigate negative implications of unsatisfactory performance.

- **Retirement Phase** (not shown in Figure 3.1): Retirement planning aims for adherence to retention policies. Characteristic for AI sub-systems is the management of models and ensuring the traceability of their model input and model output.

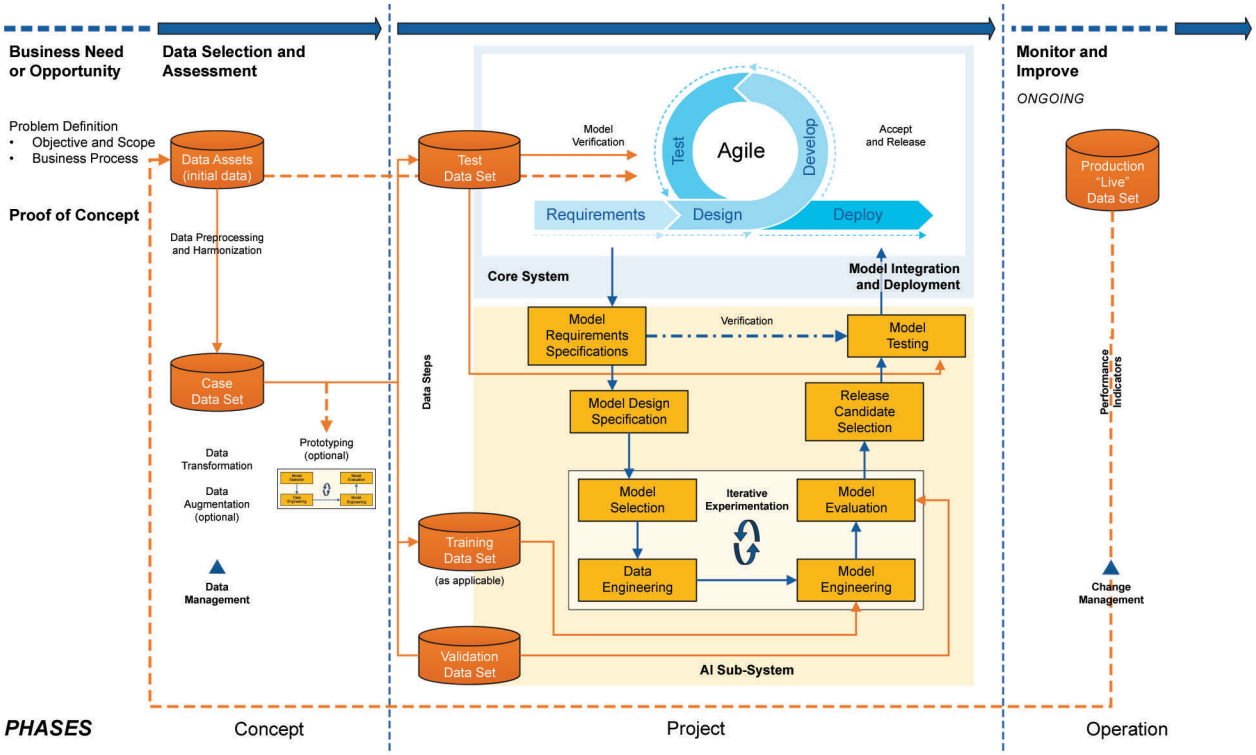
At the core of this life cycle model is a model development life cycle, which is characterized by an iterative approach to derive models. The model design is specified, which describes the range of applicable models. Models include pre- and post-processing functionality and are derived via an iterative experimentation loop; see Section 2.4.4 for further details:

1. Model selection
2. Data engineering
3. Model engineering
4. Model evaluation

Out of this set of models, suitable release candidates are selected that adhere to the model design.
Model development activities may occur at various phases:

- **Prototyping:** Establishing a simpler, though running version, typically during concept phase to gain insights into the feasibility of the use of AI; see Appendix P1
- **Development of the AI sub-system:** Developing the AI sub-system, including its models, with the goal of integration into the AI-enabled computerized system in the project phase; see Appendix P2
- **During operation of the AI sub-system:** Deriving iterations, which may be as simple as retraining with new data, up to more complex model redesign activities; see Appendix P3.

Figure 3.1: Life Cycle Model including an AI Sub-System



3.3 Specification and Verification

Specification activities include upfront determination of intended systems, components, or functionality, while verification ensures that specifications are met. Typically, specification and verification activities are organized across various levels, where a specification activity maps to a verification activity; see Section 4.3.7.

As presented in *ISPE GAMP 5 (Second Edition)* [2], linear and incremental ("agile") approaches may be followed, both compatible with the specification and verification approach.

For AI sub-systems, a specification and verification approach also applies; model requirements specifications are verified by means of model testing. However, the model evaluation during iterative experimentation is not seen as a verification activity, as it primarily guides the iterative development approach; see Appendix P2 for details.

3.4 AI-Enabled Computerized System Validation Framework and Critical Thinking

> *"GAMP advocates a computerized system validation framework to achieve and maintain GxP compliance throughout the computerized system life cycle."* [2]

While fundamental framework elements such as the use of system-specific validation plans and reports and the application of appropriate controls are applicable in the context of AI, additional elements in the validation framework are needed to achieve effective and high-quality AI-enabled computerized systems in line with regulatory expectations.

The complexity and the potential probabilistic nature of models, and AI sub-systems, leads to a high reliance on performance indicators for the determination of their fitness for purpose. In turn, the determination of fitness for intended use of their AI-enabled computerized systems and the validated state depends on these measures.

As for non-AI enabled computerized systems, decisions should be based on the level of risk when using AI. These decisions should consider AI-specific risks that emerge from the interplay of data and models and their impacts on the processes they support, and the possibilities to improve AI-enabled computerized systems as new data is created. Based on risks and applying critical thinking, regulated companies should consider the scalability of life cycle activities as part of the AI-enabled computerized system validation framework.

> *"For automated manufacturing equipment [using AI], separate computer system validation should be avoided. Computer system specification and verification should be part of an integrated engineering approach to ensure compliance and fitness for intended use of the complete automated equipment. Further information can be found in ISPE Baseline® Guide: Volume 5 – Commissioning and Qualification (Second Edition)."* [2]

Further information on risk-based approaches and critical thinking to validation is located in Chapter 5 and Appendices M3 and M4.

AI Governance underpins the safe and effective use of AI concerned with organizational structures, policies, and processes. Regulated companies should incorporate such governance considerations in their AI-enabled computerized system validation framework, which also supports adherence to trustworthy AI principles; see Appendix M9.

4 Life Cycle Phases

4.1 Introduction

This chapter provides more detailed information on typical stages and activities within the life cycle approach phases introduced in Chapter 3, and the integration of activities concerned with the AI-enabled computerized system, the AI sub-system, and models.

Guidance presented here is not prescriptive; regulated companies and other organizations, such as suppliers, should fit their life cycle and development processes to their organizational structures and business focus.

While the phases are sequentially presented, revisiting an earlier phase is a valid option depending on the situation. According to their interpretation of the life cycle model, regulated companies should record and review such events to improve processes and decision-making in the future.

4.2 Concept

The goal of the concept phase is to derive an informed decision on whether to address a business need or opportunity via development activities in the project phase. Activities are conducted in this phase to derive initial requirements and further insights as the foundation for subsequent phase activities.

Planning a PoC occurs to structure subsequent steps and capture information as information is gained, eventually allowing for acceptance (or rejection) of the PoC.

Suppliers may be involved in a variety of activities, see Appendix P1.

4.2.1 Initiation

The life cycle is initiated by identification of a business need or an opportunity, including the potential for improvements requested by stakeholders or evident from data, or from unsatisfactory performance of a live AI-enabled computerized system. The first goal is to derive a problem statement and a use case definition that should be assessed by subsequent activities.

Stakeholder engagement is required for the initiation of life cycle activities. It includes planning of communication that aims for ongoing stakeholder engagement throughout the life cycle.

See Sections 8.4 and 8.5 for further details.

4.2.2 Solution Ideation

The problem statement and the use case definition are addressed by generating possible solution approaches, leveraging stakeholder engagement. It is during the selecting and prioritizing of these approaches that the use of AI may be identified as a potential option.

See Section 8.6 for further details.

4.2.3 Context of Use, Scope Definition, and Initial Draft Requirements

The context of use and its scope covered by the solution are established to capture the potential benefits for the process. Initial draft requirements capture expectations that inform subsequent steps, such as implementing a prototype, and contribute to deciding on the success or failure of the concept phase.
See Sections 8.7 and 8.8 for further details.

4.2.4 Initial Risk Assessment and Trustworthy AI Considerations

An initial risk assessment is performed (or revisited if a prior version is in place) to determine the GxP relevance and impact. Initial considerations on potential hazards or controls are captured. As part of the assessment, alignment with trustworthy AI principles should be referenced to inform further requirements and design choices in subsequent activities.

See Section 8.9 and Appendix M9 for further details.

4.2.5 Feasibility Assessment

A feasibility assessment informs the PoC by considering the availability of data, aiming for support of a solution based on AI. Development of the initial data understanding assists in determining the suitability and possible shortcomings of data in the context of use of the model.

A selection of possible model types, or existing models, is established that could address the business need or opportunity, and the problem statement.

An optional element in the concept phase is implementing a prototype as a simplified version of the model development life cycle to derive first insights into its potential and possible limitations. A prototype is recommended in case of limited experience in use of AI for the use case.

See Section 8.10 for further details.

4.2.6 Evaluation and Acceptance of the Proof of Concept

Results from previous activities in the concept phase are summarized to seek alignment across stakeholders and derive a management decision of whether the project phase should be initiated.

See Section 8.11 for further details.

4.2.7 Capture of Initial Requirements

Initial requirements are captured for the start of project phase activities, structuring the information gathered throughout concept phase activities. They describe the intended AI-enabled computerized system, or its new or revised AI sub-system(s) in the case of an existing system.

See Section 8.12 for further details, as well as *ISPE GAMP 5 (Second Edition)* Appendix D1 [2].

4.3 Project

The project phase covers development activities from planning to releasing the AI-enabled computerized system. Development includes developing AI and non-AI functionality following various steps of specification and verification.

4.3.1 Planning

> *"Planning should cover all required activities, responsibilities, procedures, and timelines."* [2]

Of key importance of planning activities is the scaling of life cycle activities. Scaling should be based on the system impact and understanding of risks, the system's complexity and novelty, and supplier capabilities. [2]

Initial requirements captured in the concept phase and insights from prototypes, if applicable, form the basis for planning activities.

Specific to the use of AI, planning includes:

- Data management activities, including selection of data and preparation of data for testing purposes
- Establishing acceptance criteria for the AI sub-system and for user acceptance testing of the AI-enabled computerized system

These inform—in addition to general planning considerations—the validation plan and thus the determination of fitness for intended use of the AI-enabled computerized system later in the project phase (see related aspects for validation reporting in Section 4.3.5).

Products and services may be integrated, contributing to development activities or becoming a part of the AI-enabled computerized system.

See Chapter 5 and Appendices M3, M4, M5, M8, and P2, as well as *ISPE GAMP 5 (Second Edition)* Appendix M1 [2] for further details.

4.3.2 Specification and Design

> *"The role of specification is to enable systems to be developed, verified, and maintained. The number and level of detail of the specifications will vary depending upon the type of system and its intended use."* [2]

Specific to AI sub-system is establishing model requirements specifications to capture expectations on the performance of models, thus guiding the model development process.

Requirements and design specifications determine technical implementation aspects, such as the creation of the case data set used for model development or the selection of a specific model design.

The integration and interrelation of specification and design aspects of models within their context of use, and their integration into the AI-enabled computerized system, needs to be considered, aiming for seamless integration of the resulting AI sub-system.

Specification and design aspects may be combined or expanded depending on the complexity of the planned AI sub-system, the complexity of the AI-enabled computerized system, and the experience of the organization.

Section 4.3.7 provides practical examples of the possible layout of specifications steps and their relation to verification steps (see Section 4.3.4).

See Appendix P2; further guidance is provided in *ISPE GAMP 5 (Second Edition)* Appendices D1 and D3 [2].

4.3.3 Model Development

Model development includes establishing model design specifications to capture applicable model types. These determine the range of models applicable for iterative experimentation. See Section 2.4.4.

Iterative experimentation relies on a suitable case data set that is fit for purpose and data splitting.

The goal is to identify a suitable model release candidate, or a smaller selection of such models with a high likelihood of meeting model requirements specifications. Models and their specific designs are typically captured by tools. Model release candidates should be consistent with the model design specification.

Iterative experimentation typically is performed in parallel to the implementation of non-AI related functionality, for example, user interfaces or integration; see Section 4.3.7.

Note: Iterative experimentation may be performed by parties other than the regulated company. However, insights derived from model development activities into the possibilities and limitations are still informative for the regulated company to determine control strategies and testing approaches.

These activities are associated with the AI sub-system layer; see Appendices P2 and M8 for further details.

4.3.4 Verification

> *"Verification confirms that specifications have been met. This may involve multiple stages of reviews and testing depending on the type of system, the development method applied, and its use."* [2]

Section 4.3.7 provides practical examples of the possible layout of verification steps and their relation to specification steps.

Verification activities include model testing to verify the performance compared to model requirements specifications. It is important that model testing is performed on data that is a) independent from previous iterative experimentation activities to the extent possible, and b) meets its fitness for purpose in the context of use defined by the regulated companies.

In addition, verification activities include the integration of AI sub-systems including models into the AI-enabled computerized system and requirements testing on the business process level the system intends to support.

Specific aspects in the testing of AI-enabled computerized systems include the human-AI interaction, such as the effectiveness of XAI methods, as well as change management aspects concerned with establishing new model versions.

Regulated companies should scale such testing activities based on the level of risk.

See Appendices P1, M1, and M3. *ISPE GAMP 5 (Second Edition)* Appendices D5 and D3 [2] provide further details.

4.3.5 Reporting and Release

> *"The system should be accepted for use in the operating environment and released into that environment in accordance with a controlled and documented process."*
>
> *"A computerized system validation report should be produced summarizing the activities performed, any deviations from the plan, any outstanding and corrective actions, and providing a statement of fitness for intended use of the system."* [2]

Of particular interest for AI-enabled computerized systems is the summarization of the insights on iterative experimentation and model testing results as well as the AI-specific aspects of user acceptance testing. This typically includes the determination of whether data used for testing is fit for purpose, and a description of how far model requirements specifications were fulfilled during model testing (see also related planning aspects in Section 4.3.1).

A summary of insights into the possibilities and limitations of models gained during iterative experimentation may be included, as they may influence activities in the operation phase and its supporting processes, such as training, ongoing monitoring, and incident and problem management.

The reporting and release marks the transition to the operation phase, which is initiated by handover activities.

See Appendices P2 and P3 and *ISPE GAMP 5 (Second Edition)* Appendices D5, M7, and O1 [2].

4.3.6 Supporting Processes

4.3.6.1 Risk Management

"An appropriate risk-management process should be established." [2] In particular, functional risk assessments are performed and controls identified that inform the design of models and their AI sub-systems. Regulated companies should integrate both AI-related and non-AI-related risk management activities into an overall risk management strategy.

See Appendices P2 and M3 for specific risk management activities in the context of AI and *ISPE GAMP 5 (Second Edition)* Appendix M3 [2] for general risk management activities.

4.3.6.2 Project Change and Configuration Management

> *"Appropriate configuration management processes should be established such that a computerized system and all its constituent components can be identified and defined at any point during its life cycle." [2]*

Regulated companies may consider software products, or a selection of tools, to support project change and configuration management (Machine Learning Operations (MLOps)).

See Appendices P2 and M8 and *ISPE GAMP 5 (Second Edition)* Appendix M8 for further details.

4.3.6.3 Design Review

> *"At suitable stages during the life cycle, planned and systematic design reviews of specifications, design, and development should be performed. This design review process should evaluate deliverables to ensure that they satisfy the specified requirements."* [2]

Of particular interest for AI-enabled computerized systems is the design review of models; this includes the choice of model types within the model design space, the architecture of individual models, considering model alternatives that may not be known at the time of planning.

> *"The rigor of the design review process and the extent of documentation should be based on risk, complexity, and novelty. Design reviews are mainly applicable for custom applications and within the supplier product development processes." [2]*

See Appendices P2 and *ISPE GAMP 5 (Second Edition)* Appendix M5 [2] for further details.

4.3.6.4 Traceability

See *ISPE GAMP 5 (Second Edition)* Appendix M5 [2] for guidance on traceability of business process needs, requirements, functional and design elements in specifications, and results of verification activities.

4.3.6.5 Documentation and Knowledge Management

> *"Management of documentation includes preparation, review, approval, issue, change, withdrawal, and storage. …Information and records may be maintained in…tools rather than in traditional documents."* [2]

IT infrastructure elements and tools may support capturing information and records established throughout the project phase and knowledge handover at the closure of the project phase when transitioning to the operation phase.

See Appendices P2, M5, and M8, Section 10.3, and *ISPE GAMP 5 (Second Edition)* Appendix M9 [2].

4.3.7 Practical Examples

Computerized systems are generally made up of a combination of components from different software categories [2], (see Appendix M11). The same applies to AI-enabled computerized systems. These include:

- Category 3 – Standard Products
- Category 4 - Configured Products
- Category 5 – Custom Applications

These categories should be understood as a continuum [2].

The software category is one factor to consider in a risk-based approach. Life cycle activities should be scaled based on the GxP impact, and the complexity and novelty of the system. [2]

For convenience, activities in this section are shown matching the representation in *ISPE GAMP 5 (Second Edition)* Section 4.2.6 [2], adding aspects on the specification and verification of models for Category 3, Category 4, and Category 5. Typically, development activities for AI-enabled computerized systems follow more agile approaches.

Examples are intended to be indicative and for illustration purposes. Actual approaches for specific systems should be based on the results of supplier assessments and risk assessments, applying critical thinking, in addition to the categorization of system components in *ISPE GAMP 5 (Second Edition)* and the considerations provided in this ISPE AI Guide in Appendices M2, M3, and M4.

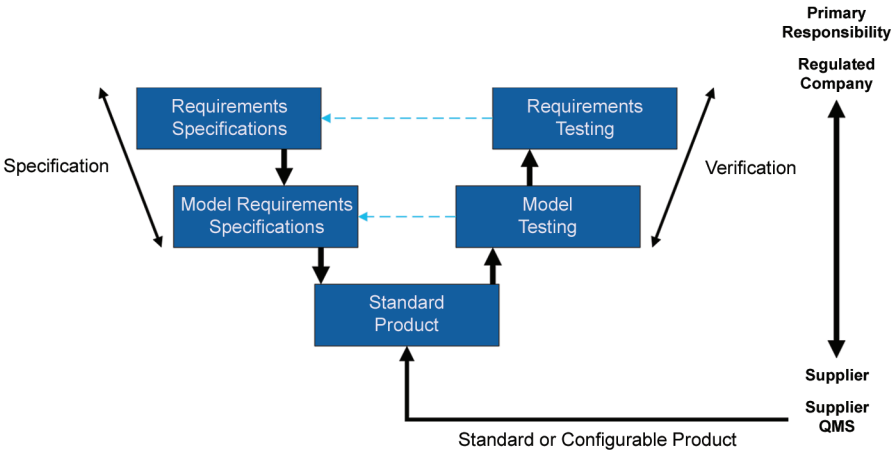
4.3.7.1 Example of a Standard Product

> *"Many computerized systems comprise commercially available software products running on standard hardware components."* [2]

Software products that are used off-the-shelf are typically classified as GAMP Category 3, including those used for business purposes [2]. An example of an AI-enabled computerized system is the use of LLMs provided in a software product to support staff with a chat bot in a secure environment, allowing no configurability or containing defined parameterization ranges.

In such cases, regulated companies may choose an approach using two levels of specification and verification, including requirements specifications and model requirements specifications and requirements testing and model testing, respectively; see Figure 4.1. Compared to the simpler approach shown in *ISPE GAMP 5 (Second Edition)* [2], the additional verification step of model requirements specifications determines the fitness for purpose of the model within the context of use of the regulated company.

Figure 4.1: Approach for a Standard Product



Testing typically covers:

- Correct installation
- Testing model requirements specifications using suitable data
- Tests that demonstrate fitness for intended use

Regulated companies typically perform requirement testing and may also perform model testing. They may distinguish data that is used for model testing (larger) and requirements testing (smaller).

Supplier activities typically include supply of the product, and provision of documentation, training, and support and maintenance, similar to non-AI-enabled configured products [2]. They may also support model testing activities leveraging test data that is fit for purpose for the regulated company's context of use.

4.3.7.2 Example of a Configured Product

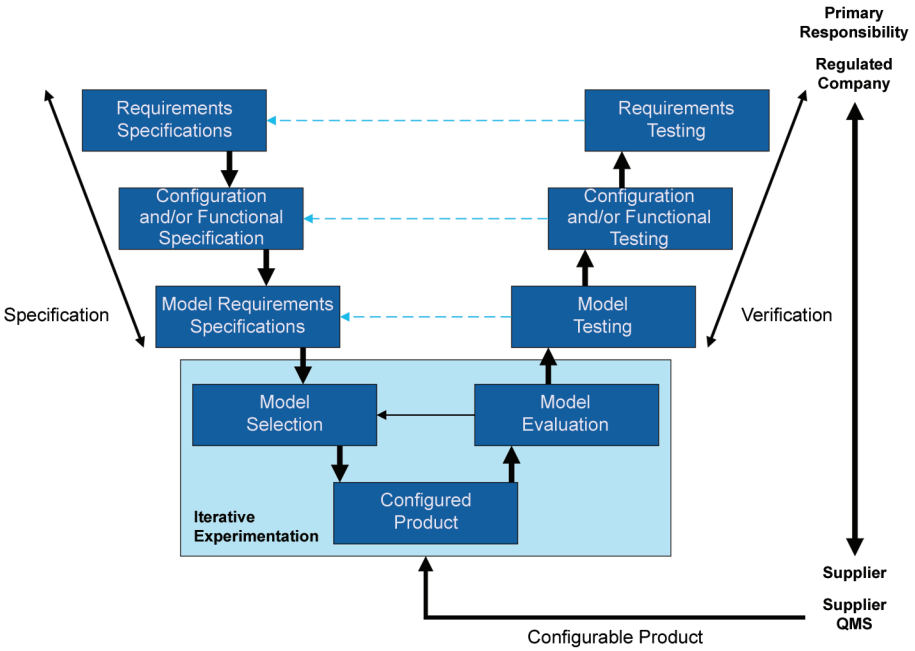
"A common type of computerized system involves the configuration of commercially available software products running on standard hardware components. …Commercial software products that are configured for a specific business process are typically classified as GAMP Category 4." [2] Similar AI-enabled products include those that integrate parts of the organization's information or allow for custom adjustment of prompts when using complex LLMs.

"In such cases, and based on satisfactory supplier and risk assessments, a flexible approach to specification and verification may be applied. The number of deliverables required to cover the required specification and verification of functionality delivered via configuration will depend on the size, complexity, and technical architecture of the system" [2], as well as the model complexity and the way that interaction with the model is configured.

"For example, in many cases, functionality and configuration aspects may be combined into one specification." [2] Similarly, as the model may be given by the product, the model design and model selection may be combined.

Figure 4.2 shows project phase activities from the regulated company's perspective. In addition to the configuration and/or functional specification and testing foreseen in this case, activities are planned to verify model requirements specifications upon deriving a suitable model during iterative experimentation within the model design space offered by the configurable product.

Figure 4.2: Approach for a Configured Product



"Testing typically covers:

- *Correct installation*
- *Configuration of the system*
- *Functionality that supports the specific business process based on risk and supplier assessments"* [2]
- Models, during iterative experimentation and as model testing upon selection of a release candidate
- Further tests resulting from risk and supplier assessments [2]

"Regulated companies should decide upon the required levels of specification and verification, and many of the project phase activities and documents may be delegated." [2] However, they should determine the suitability of data used for model testing that is deemed to provide sufficient evidence of its fitness for purpose in their context of use. As for non-AI-enabled computerized systems, the number of deliverables required for specification and verification depends on the size, complexity, and technical architecture of the system, and the complexity of embedded models.

"Supplier activities typically include:

- *Supply of the product [including embedded models]*
- *Production of specifications and test specifications, as required, on behalf of the regulated company*
- *Support during configuration and testing"* [2]
- Support during iterative experimentation
- *"User documentation*

- *Training*
- *Support and maintenance activities"* [2]

Support for model development activities, the product, and its configuration may be performed by different suppliers.

4.3.7.3 Example of a Custom Application

"Some computerized systems are developed to meet individual user requirements, where no commercially available solution is suitable. …The software developed for such systems is classified as GAMP Category 5." [2]. Similar considerations apply to AI-enabled computerized systems in which custom models are used. Examples include implementation of custom models, training of sourced models based on the regulated company's data sources, and fine-tuning of provided models with own data as an element of a custom application.

When developing such custom models, various levels of specification and verification may be applied, depending on whether those models are implemented within a software product, or if a fully custom application is built. Figure 4.3 shows an example approach for a fully custom application. This approach consists of groups in two layers, with interactions between them:

1. AI-enabled computerized system layer: Typical specification and verification activities for requirements, functions, designs, and modules, capturing the non-AI components and the overall integration in the AI-enabled computerized system
2. AI sub-system layer: Typical specification and verification activities for the AI sub-system, and model design and model selection.

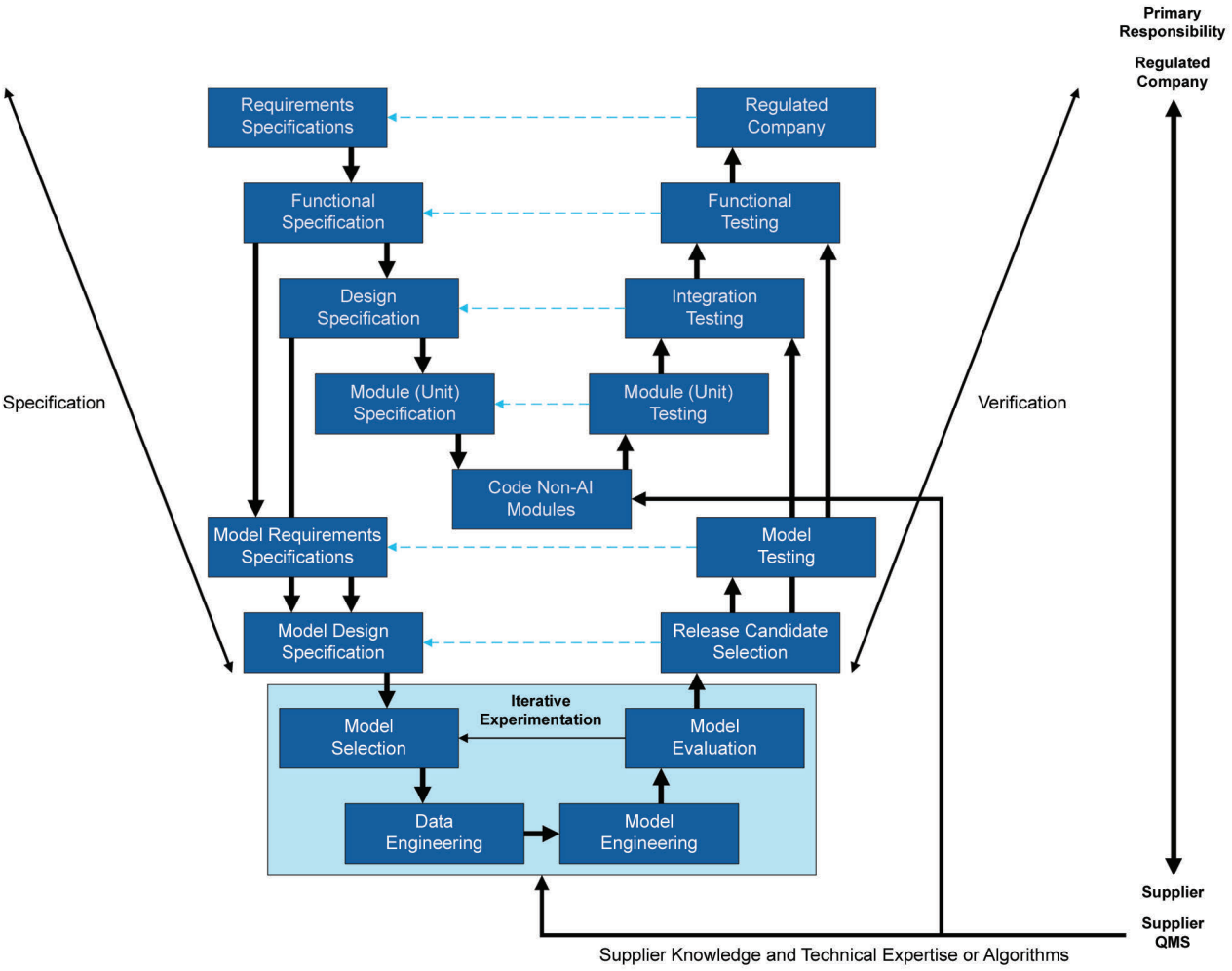
Coding of non-AI modules, if not part of data engineering during iterative experimentation, is assumed to also cover integration of models within the AI sub-system.

Links between AI-enabled computerized systems and the AI sub-system and models include:

- Functional specifications that inform model requirements specifications
- Design specification that informs the model design specification
- Release candidates chosen for Integration testing
- Verified AI sub-system integrated for functional testing

Development activities include the full iterative experimentation process.

Figure 4.3: Approach for a Custom Product



"Testing typically covers:

- *Correct installation*
- *Functionality and design"* [2]
- Model requirements specifications
- Models, during iterative experimentation and upon selection a release candidate
- *"Tests that demonstrate fitness for intended use*
- *Any further tests as a result of risk assessments and supplier assessments"*

"Regulated companies should decide upon the required levels of specification and verification, and many of the project phase activities and deliverables may be delegated." [2] Since the system is new, and new AI sub-systems are established, rigorous testing should be performed at functional and design levels, as well as on AI sub-systems.

Supplier activities typically include production of specifications and test specifications on behalf of the regulated company, development of new software, testing, user documentation, training, and support and maintenance activities. [2] In addition, suppliers may support model design activities, conduct iterative experimentation, as well as select release candidates and perform model testing.

> *"Complex systems may require a further hierarchy of specifications covering hardware design specifications and configuration specifications."* [2]

4.4 Operation

The operation phase aims to maintain the safety and efficacy of the AI-enabled computerized system while allowing incremental changes, including iterations of the AI sub-systems and models. Suppliers typically maintain their products as well as support continual improvement. Various supporting processes occur throughout the operation of an AI-enabled computerized system.

> *"The use of good IT practices (e.g., ITIL…), supported by IT service management tools and automation is encouraged to ensure the efficiency and effectiveness of support processes and an auditable record of support activities."* [2]

For more information on system operation topics, see the *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems* [57].

Operational processes *"are supported by QMS activities, such as document management, records management, knowledge management, training management, and the maintenance of up-to-date end user procedures."* [2] Further information is available in *ISPE GAMP® Guide: Records and Data Integrity* [15] and *ISPE Good Practice Guide: Knowledge Management in the Pharmaceutical Industry* [38].

4.4.1 Handover

Handover summarizes the activities that are required to transition the AI-enabled computerized system after successful release to stable operation. Knowledge transfer is of high importance, including relevant insights gained during the development process, as well as verification or testing activities.

In addition to non-AI-enabled computerized system handover activities, knowledge transfer should cover insights on the interplay of data and models, and their limitations, to support effective and safe operation of the AI-enabled computerized system.

See Section 10.3 and *ISPE GAMP 5 (Second Edition)* Appendix O1 [2] for further details.

4.4.2 Service Management and Performance Monitoring

"The support required for each system, and how it will be provided, should be established." [2] In addition to non-AI-enabled computerized systems support services, the system support should cover the ongoing monitoring of the AI sub-systems' and their models' performance based on a clear understanding of the relationship between performance indicators, risks, and risk controls.

See Section 10.4 and *ISPE GAMP 5 (Second Edition)* Appendices O2, O3, and D9 [2].

4.4.3 Incident and Problem Management and Corrective and Preventive Action

4.4.3.1 Incident Management and Problem Management

> *"The incident management process aims to categorize incidents to direct them to the most appropriate resource or complementary process to achieve a timely resolution; whereas problem management involves analyzing root causes and preventing incidents from happening in the future."* [2]

Incident and problem management for AI-enabled computerized systems should consider AI-specific incidents, such as the occurrence of bias or drifts in data and models or ineffective human-AI interaction. While general types of incidents may serve as a starting point, regulated companies should elaborate on possible incidents in the specific context of use, based on product and process understanding and critical thinking.

See Section 10.6 and *ISPE GAMP 5 (Second Edition)* Appendix O4 [2].

4.4.3.2 Corrective and Preventive Action (CAPA)

> *"CAPA is a process for investigating, understanding, and correcting discrepancies based on root-cause analysis, while attempting to prevent their recurrence."* [2]

Since the use of AI may lead to specific types of incidents, further corrections and prevention may apply in addition to incidents from non-AI-enabled computerized systems. Regulated companies should maintain information and records of CAPA processes, while assessing their effectiveness.

See Section 10.7 and *ISPE GAMP 5 (Second Edition)* Appendix O5 [2].

4.4.4 Change Management

4.4.4.1 Change Management

> *"Change management is a critical activity that is fundamental to maintaining the proper functioning and controlled status of systems and processes. …The process should allow the rigor of the approach, including the extent of documentation and verification, to be scaled based on the nature, risk, and complexity of the change, by application of critical thinking."* [2]

Specific to AI-enabled computerized systems, the management of model iterations is of high relevance. This includes establishing new model versions, according to static or dynamic system design choices.

Some changes such as the change of model types may lead to a substantial redesign, thus suggesting activities as those outlined in the concept phase or project phase.

See Section 10.8 and *ISPE GAMP 5 (Second Edition)* Appendix O6 [2].

4.4.4.2 Configuration Management

> *"Configuration management includes those activities necessary to precisely define a computerized system at any point during its life cycle, from the initial steps of development through to retirement."* [2]

In the context of AI-enabled computerized systems, the implications of configurations, e.g., controlling the use of data or data transformations, or instructing and controlling models, should be carefully assessed, as the model behavior may change substantially because of such changes. Separate model testing activities may be considered, based on the risk.

See Section 10.8 and *ISPE GAMP 5 (Second Edition)* Appendix O6.

4.4.4.3 Repair Activity

> *"The repair or replacement of defective computerized system components…should be managed in accordance with a defined process. …Many repair activities are emergencies and require rapid resolution."* [2]

While some activities to maintain AI sub-systems as part of AI-enabled computerized systems may exhibit similar patterns to repair activities (e.g., the correction of data used for monitoring), the impact of repair activities on hardware or infrastructure components should be considered. Of similar nature are patches and updates of supporting infrastructure, for example, the compatibility of the model in its runtime environment may be affected by server updates.

See *ISPE GAMP 5 (Second Edition)* Appendix O6 [2] for further details.

4.4.5 Periodic Review

> *"Periodic reviews are used throughout the operational life of systems to verify that they remain compliant with regulatory requirements."* [2]

Specific to AI-enabled computerized systems is the periodic review of risks, data and data changes, model performance, changes of models, insights from incident and problem management, and potential changes in statutes and regulations.

See Section 10.9 and *ISPE GAMP 5 (Second Edition)* Appendix O8 [2].

4.4.6 Continuity Management

Per *ISPE GAMP 5 (Second Edition)* [2], regulated companies should establish the following processes for continuity management:

- Backup and Restore: Ensure that backup copies of software, records, and data are made, maintained, and retained for a defined period within safe and secure areas.
- Business Continuity Planning: Ensure the organization is fully prepared to respond effectively in the event of failures and disruptions, covering local and global infrastructure, data, and the application.
- Disaster Recovery Planning: Plans for the recovery of specific systems in the event of disaster.

Aspects of continuity management specific to AI-enabled computerized systems include quality of data, integrity of AI sub-systems and models, as well as potential impact on performance metrics, all of which have implications on the system's state of control. Robust data and model governance and management practices help mitigate negative impacts and support quick uptake of safe and effective operations.

See Section 10.10 and *ISPE GAMP 5 (Second Edition)* Appendices O9 and O10 for further details.

4.4.7 Security and System Administration

> *"Computerized systems and data should be adequately protected against willful or accidental loss, damage, or unauthorized change."* [2]

AI-specific cybersecurity risks are of particular importance for AI-enabled computerized systems, including potential negative implications on models, and the possibility to extract sensitive information via exposed functionality. Regulated companies should safeguard against and detect possible attacks based on an understanding of the AI-enabled computerized system and its model(s) vulnerabilities.

System administration should be performed in line with *ISPE GAMP 5 (Second Edition)* [2].

See Appendix S5 and *ISPE GAMP 5 (Second Edition)* Appendices O11 and O12 for further details.

4.4.8 Record Management

"Policies for the retention of regulated records should be established, based on a clear understanding of regulatory requirements and existing corporate policies, procedures, standards, and guidelines."

"Archiving is the process of taking records and data off-line by moving them to a different location or system, often protecting them against further changes." [2]

In addition to guidance in *ISPE GAMP 5 (Second Edition)* [2], the traceability of critical data between model input, the model, and its output should be maintained to allow for *ex post* assessment.

See *ISPE GAMP 5 (Second Edition)* Appendices O11 and O12 for further details.

4.5 Retirement

System retirement terminates the life cycle, which means that no productive use is foreseen for the AI-enabled computerized system or the retired sub-system unless reactivated. Activities include planning, execution, and reporting captured records and information.

Data and models, including the traceability of model input, the model, and model output, are of primary importance when retiring AI-enabled computerized systems to allow for *ex post* assessment, when needed.

System retirement also provides an opportunity for stakeholders to review the life cycle and identify areas for improvement.

See Appendix P4, *ISPE GAMP 5 (Second Edition)* Appendix M10 [2], *ISPE GAMP Guide: Records and Data Integrity* [15], and *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* [1].

5 Quality Risk Management (QRM)

5.1 Introduction

The concept of QRM is introduced in Chapter 2 as part of the life cycle approach. This chapter includes details from *ISPE GAMP 5 (Second Edition)* [2] and incorporates factors related to AI. While the core principles of QRM are the same as those for non-AI enabled computerized systems, the unique characteristics of AI-enabled computerized systems introduce new aspects.

This chapter is primarily aimed at newly implemented AI-enabled computerized systems, although it also informs the extension of existing computerized systems with AI sub-systems.

> *"It does not imply that formal risk assessments are required for all existing systems. The extent of risk management required for existing systems, including the need for formal risk assessments, should be considered as part of periodic review."* [2]

See Appendix M3 and *ISPE GAMP 5 (Second Edition)* Appendix M3 for further details.

5.2 Overview

"QRM is a systematic process for the assessment, control, communication, and review of risks." [2] This iterative process should be used throughout the AI-enabled computerized system life cycle from concept to retirement.

> *"For a given organization, a framework for making risk-management decisions should be defined to ensure consistency of application across systems and business functions."*
>
> *"Such a framework is most effectively implemented when it is incorporated into the overall QMS and is fully integrated with the system life cycle."* [2]

5.3 Science-Based QRM

Per *ISPE GAMP 5 (Second Edition)* [2]:

> *"Determining the risks posed by a computerized system requires a common and shared understanding of:*
>
> - *Impact of the computerized system on patient safety, product quality, and data integrity*
> - *Supported business processes*
> - *[Critical Quality Attributes] CQAs for systems that monitor or control CPPs [Critical Process Parameters]*
> - *User requirements*
> - *Regulatory requirements*
> - *Project approach (contracts, methods, timelines)*
> - *System components and architecture*
> - *System functions*

- *Supplier capability*

The organization also should consider other applicable risks such as Health, Safety, and Environment (HSE)."

In the context of AI-enabled computerized systems, additional risk factors include:

- Shortcomings in data quality and their possible implication on model performance
- Inadequate choice of models, which may lead to inferior performance or bias
- Occurrence of model drifts or data drifts during operation
- Insufficient human-AI interaction, reducing effectiveness of human oversight

As stated in ISPE GAMP 5 (Second Edition) [2], *"Managing risks may be achieved by:*

- *Elimination by design*
- *Reduction to an acceptable level*
- *Verification to demonstrate that risks are managed to an acceptable level"* [2]

Risks should be eliminated through process or system design where possible. Design reviews can help to achieve this.

For risks that cannot be eliminated, they should be reduced to an acceptable level through controls or procedures that aim to lower severity, decrease probability, or increase detectability.

"A systematic approach should be defined to verify that the risk associated with a system has been managed to an acceptable level. The overall extent of verification and the level of detail of documentation should be based on the risk to patient safety, product quality, and data integrity, and take into account the complexity and novelty of the system." [2]

Risk information may emerge at different stages of the life cycle and should be considered accordingly. *"For example, the high-level risks associated with a business process need to be understood before the risks associated with specific functions of computerized systems can be assessed."*[10]

"The criticality of a business process is independent of whether it is manually processed, semi-automated, or fully automated. [Computerized] systems that support critical processes include those that:

- *Generate, manipulate, or control data supporting regulatory safety and efficacy submissions*
- *Control critical parameters and data in pre-clinical, clinical, development, and manufacturing*
- *Control or provide data or information for product release*
- *Control data or information required in case of product recall*
- *Control adverse event or complaint recording or reporting*
- *Support pharmacovigilance"* [2]

[10] Per *ISPE GAMP 5 (Second Edition)* [2]: *"CQAs of drug development and manufacture will influence understanding of the impact of business process, while CPPs will influence the impact of specific computer functions."*

5.4 QRM Process

Per *ISPE GAMP 5 (Second Edition)* [2]:

> *"ICH Q9…describes a systematic approach to QRM intended for general application within the pharmaceutical industry. It defines the following two primary principles of QRM:*
>
> - *'The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and*
> - *The level of formality and documentation of the quality risk-management process should be commensurate with the level of risk.'*
>
> *In the context of computerized systems, scientific knowledge is based upon the system specifications and the business process being supported."*

This Guide uses the following key terms from ICH Q9(R1) [31]:

> *"**Harm:** Damage to health, including the damage that can occur from loss of product quality or availability.*
>
> ***Hazard:** The potential source of harm.*
>
> ***Risk:** The combination of the probability of occurrence of harm and the severity of that harm.*
>
> ***Severity:** A measure of the possible consequences of a hazard."*

This Guide applies the general principles of ICH [13] to describe a five-step process for risk management as an integral part of achieving and maintaining system compliance. For simple or low-risk systems, some of these steps can be combined. See Appendix M3 for further details on the QRM process.

Risk management should be applied throughout all phases of the life cycle.

ISPE GAMP 5 (Second Edition) is focused on managing risks during the project phase and highlights the appropriate use of risk management during the operational phase. Examples include:

- Determining the need for supplier audit as part of supplier assessment
- Determining the rigor and extent of testing
- Determining the corrective actions arising from test failures
- Determining the impact of changes as a part of change management
- Determining the frequency of periodic reviews

For AI, risk management should consider the following aspects and implications throughout the life cycle:

- Implications of data quality on the performance of models
- Complex models can exhibit "black box" characteristics, i.e., their complexity does not admit meaningful assessment of the exact mechanics by which a model derives its model output from model input; hence, selection of models imply impact on patient safety, product quality, and data integrity

- The evaluation of models relies on suitable upfront decided metrics and thresholds that require a connection between risk management activities and measurement of model performance

- ML models offer the option of dynamic system design, i.e., adaptation as new data is collected; dynamic systems introduce further risks, such as model drift, and require respective controls

"Organizations may have established risk-management processes, including the use of methods such as those listed in [ISPE GAMP 5 (Second Edition) or those listed in Appendix M3 of this ISPE AI Guide]. While this Guide describes one suggested approach, it does not intend or imply that these existing methods should be discarded, rather that they continue to be used, as appropriate, within the context of an overall QRM framework consistent with ICH Q9." [2]

Process Risk Assessment

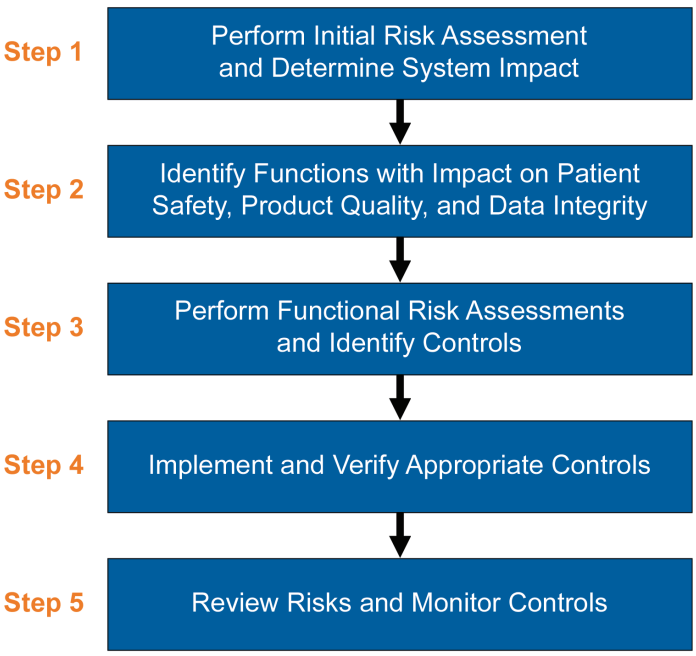
"Some records and data may reside on more than one system during their life cycle, and QRM activities should start at the business process level, at a level higher than individual systems."

"The process risk assessment is aimed at identifying key high-level risks to patient safety, product quality, and data integrity, and identifying the required controls to manage those risks. Typically, at this stage no assumptions are made about the nature or exact functionality and design of the computerized system(s) that will support the process" [2]; however, guidance in ISO/IEC 42001 [5] may help prepare for further steps and integrate management aspects in the context of using AI.

The frequency and extent of any periodic review should be based on the level of risk and should consider previous findings and operational history. [2]

The QRM process steps from *ISPE GAMP 5 (Second Edition)* [2] (shown in Figure 5.1) are listed along with the differences/additional considerations for AI-enabled computerized systems. Further information is in Appendix M3.

Figure 5.1: Quality Risk Management (QRM) Process



Step 1 – Perform Initial Risk Assessment and Determine System Impact

Regulated companies should perform an initial risk assessment of the (AI-enabled) computerized system based on business processes and business risk assessments, user requirements, regulatory requirements, and known functional areas. [2]

Determining the system impact in this step is agnostic to the use of AI, as it is concerned with the process the (AI-enabled) computerized system intends to support. In addition, first information may be collected regarding possible hazards and controls.

Step 2 – Identify Functions with Impact on Patient Safety, Product Quality, and Data Integrity

> *"Functions that have an impact on patient safety, product quality, and data integrity should be identified by building on information gathered during Step 1."* [2]

This step may be influenced by consideration of the use of AI, since the planning of AI sub-systems determines the system design and how functions are compartmentalized.

Steps 3, 4, and 5 are highly specific to AI sub-systems and the functional risks they and their embedded models introduce, including identification, implementation, and review of suitable controls; see examples below.

Step 3 – Perform Functional Risk Assessments and Identify Controls

> *"Functions identified during Step 2 should be assessed by considering possible hazards, and how the potential harm arising from these hazards may be controlled."* [2]

See Appendix M3 for examples.

Step 4 – Implement and Verify Appropriate Controls

> *"The control measures identified in Step 3 should be implemented and verified to ensure that they [are effective]."* [2]

Example control measures for AI include:

- Involvement of human verification of model results; see Appendix M10
- Choice of model designs that admit better interpretability, or use of XAI methods; see Appendix S4
- Use of a range of performance indicators that can be used for model testing and for ongoing monitoring of model performance in operation; see Appendices P2 and P3, respectively
- Use of synthetic data to expand coverage of the case data set and to test limitations of the model; see Appendices P2 and M7
- Use of a high degree of automation for handling of data and models, see Appendix M8
- Specifically for the case of Generative AI, the use of guardrails should be considered; this is a set of techniques, including control of model input and control of model output

The ML Risk and Control Framework [32] provides further examples of controls, addressing hazards throughout the AI-enabled computerized system life cycle.

Step 5 – Review Risks and Monitor Controls

> *"During periodic review of systems, [during change management], or at other defined points, [a regulated company] should review the risks. The review should verify that controls are still effective, with corrective action taken under change management if deficiencies are found."* [2]

Review and monitoring of controls specific to the use of AI relies on ongoing monitoring of data and model performance as well as user feedback and user interactions, see Appendix P3. In addition to knowledge gained through the use of the system, it should also take into account technological advancements like the availability of new models or methods that may support strengthening of the risk control strategy.

6 Regulated Company Activities

6.1 Introduction

"Responsibility for the compliance of computerized systems lies with the regulated company. This involves activities at both the organizational level and at the level of individual systems." [2]. Building on the information in *ISPE GAMP 5 (Second Edition)* [2], this section adds details for AI-enabled computerized systems to:

- Governance for achieving compliance
- System-specific activities

This chapter provides an overview of activities performed by the regulated company, although some regulated company activities refer to collaboration with a supplier, which may be internal or external to the organization. Chapter 7 describes such supplier activities.

6.2 Governance for Achieving Compliance

"Achieving robust, cost-effective compliance requires strong governance. Key elements of successful governance include:

- *Establishing computerized systems compliance policies and procedures*
- *Identifying clear roles and responsibilities*
- *Training*
- *Managing supplier relationships*
- *Maintaining a system inventory*
- *Planning for validation*
- *Continual improvement activities"* [2]

Additional elements for effective governance for AI include:

- Maintaining data and model inventory
- Establishing responsible use policies for data
- Building AI literacy
- Data ownership and data use policies
- Establishing data and model governance and management, which goes beyond data governance considerations provided in *ISPE GAMP 5 (Second Edition)* [2]

6.2.1 Computerized Systems Policies and Procedures

"Regulated companies should have a defined policy for ensuring that computerized systems are compliant and fit for intended use." [2] In addition to items listed in *ISPE GAMP 5 (Second Edition)* [2], data management practices need to be augmented by model management practices.

6.2.2 Identifying Clear Roles and Responsibilities

As stated in *ISPE GAMP 5 (Second Edition): "Roles and responsibilities for activities should be documented, allocated, and communicated. The appropriate and timely involvement of these key roles should be ensured."* [2]

The responsibilities listed in *ISPE GAMP 5 (Second Edition*) apply to AI-enabled computerized systems:

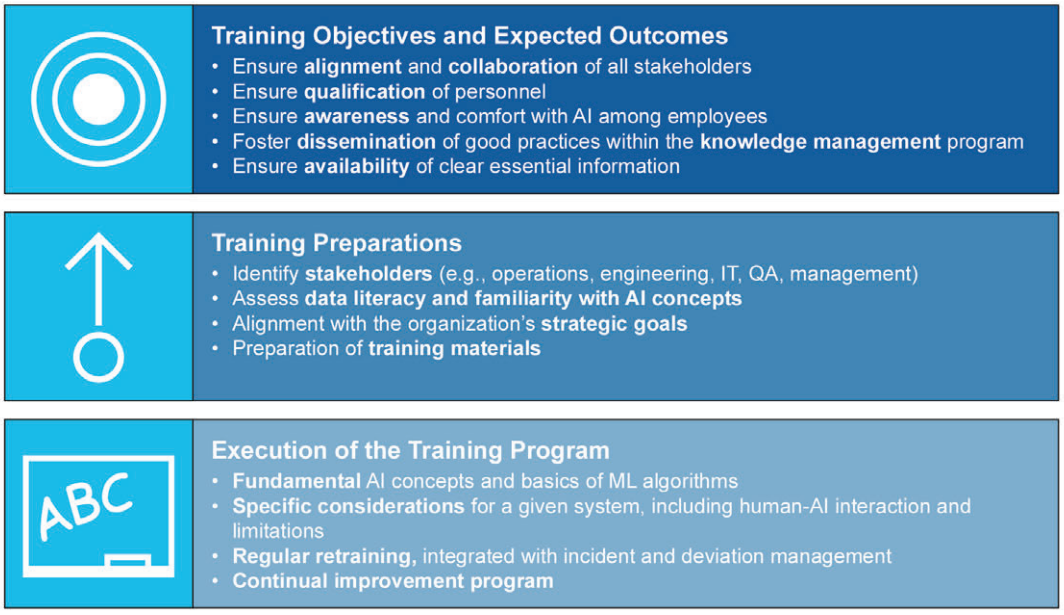
- Defining, approving, and maintaining policies and procedures
- "Compiling and prioritizing the system inventory
- Producing plans and reports
- Managing compliance and validation activities
- Maintaining compliance during operation"

While the significance of the roles such as the process owner, system owner, data owner, quality unit, SMEs, end users, and suppliers applies, they cover new aspects in the context of AI. Furthermore, regulated companies should consider new roles and responsibilities. Details are provided in Section 6.3.3.

6.2.3 Training

Education and training should be regularly executed as part of the organization's knowledge management process. Individual training on specific topics, processes, and changes falls under this process. This section highlights the goals, prerequisites, activities, and outcomes to consider. An overview is provided in Figure 6.1.

Figure 6.1: Training of Staff for AI-enabled Computerized Systems



6.2.3.1 Training Objective and Expected Outcomes

> *"Training is the process that ensures that persons who develop, validate, maintain, support, or use computerized systems have the education, training, and experience to perform their assigned tasks."* [2]

Training for AI-enabled computerized systems, specifically for their release at a site or at an organizational level, is important for ensuring that all stakeholders are aligned on the use of AI within the regulated process; see Section 6.3.3. In addition, human oversight and control is maintained to safeguard patient safety, product quality, and data integrity [33].

Key stakeholders involved in using AI directly or indirectly should be equipped with the knowledge needed and the role-specific skills to use AI appropriately in their operations. Training also fosters transparency, providing users and other stakeholders with clear and essential information [33].

Furthermore, it is beneficial to encourage a culture of continual improvement and training. This can ensure that AI is incorporated effectively in a specific area or throughout the organization, and that a broader rollout is optimized once needed.

Outcomes of such training include:

- **Material and information for training:** This includes presentation materials, usage manuals, and other resources available for reference. This material should be routinely updated to ensure it is current and available.
- **Increased awareness and comfort with AI among relevant individuals:** This can be demonstrated by their use of systems and tools in their roles, as relevant and as permitted, see also Section 6.3.3.
- **Effective collaboration around AI:** Fostering an environment of thoughtful communication in internal forums and communities of practice.
- **Dissemination of learnings to support improvements:** Improvements in areas where AI is leveraged can promote operational efficiency and reduced downtime.
- **Information to users:** *"Users are provided clear, essential information"* [33].

Example: Using Freely Available AI Models with General Web Browsers

Clear policies should be established regulating the use of free (foundational) AI models accessed through a browser in corporate office environments. Such policies should either prohibit or allow it, depending on the individual's environment and provided that a robust awareness and training program is in place. This program should highlight the risk of data privacy and Intellectual Property (IP) protection when using free AI models and should ask users to exercise caution. Additionally, training should emphasize that AI services are established with Service Level Agreements (SLAs) that guarantee privacy and confidentiality of data.

6.2.3.2 Training Prerequisites

Individuals should have obtained suitable AI literacy for the tasks they perform, aligned with organizational needs. Reaching such AI literacy should be verified.

Regulated companies should address any knowledge gaps in AI-related aspects with relevant training. A skill matrix for each role utilizing an AI-enabled computerized system is considered good practice, as is conducting individual assessments to elicit the personnel's current capabilities and to determine additional training needs.

As current processes will likely change with the introduction of AI-enabled systems and general training on relevant AI topics, specific training on the process supported with AI is also required.

6.2.3.3 Training Execution and Training Success

Training programs should be developed for organization-specific needs and requirements. These include:

- Fundamental AI concepts
- Basics of ML
- Common applications of AI and ML across the industry
- Considerations for a specific context of use:
  - General function and role of the AI sub-system
  - Focus on human-AI interactions [34]
  - Limitations that can come with the AI-enabled computerized system

Practical training including hands-on elements helps build knowledge and achieve the desired confidence level for all stakeholders. Training should also include content on data management, data quality, and regulatory compliance implications of the use of AI.

Training should not be a point in time activity; rather a continual improvement program with knowledge sharing opportunities is beneficial. Examples include stakeholder forums, workshops, and internal communities of practice dedicated to the organization-wide adoption of AI.

As the use of AI has the potential to degrade individuals' knowledge and experience over time (i.e., people are apt to rely more on the technology), "first principle" scenario-based training is recommended at regular intervals. This can be based on the risk of the process, the competency level required of an individual, and the length of time since the process was last performed.

Continual training plans should be seen as an iterative process, with input from performance monitoring, incident, and CAPA management processes as well as development activities to stay up-to-date and state-of-the-art.

6.2.4 Managing Supplier Relationships

> *"All phases of the computerized systems' life cycle require cooperation between the regulated company and external and internal suppliers, including IT and engineering. …Responsibility for activities may be with the suppliers, but in all cases regulatory accountability lies with the regulated company."* [2]

Supplier relationships in the context of AI-enabled computerized systems include:

- Considering AI-specific factors in supplier assessments and during supplier qualification
- Performing supplier risk management on supplied data, models, or AI-enabled software products
- Considerations on data privacy and IP infringement and protection
- AI-specific aspects of cybersecurity management

- Collaboration with suppliers during development of models and AI-enabled computerized systems, including choice of models and model requirements specifications
- Leveraging supplier information for verification purposes, such as the performance of models
- Operation support, considering AI-specific changes, and problem and incident management processes
- Retention of data and models after retirement

Per *ISPE GAMP 5 (Second Edition): "The regulatory expectation is that systems are fit for intended use and maintained in a state of control and compliance. Activities performed by, and information maintained by, suppliers following their own methods and approaches, and under their own QMS, may assist in achieving this objective."* [2]

Additional considerations from the supplier's perspective are provided in Chapter 7.

6.2.5 Maintaining the System Inventory

"Regulated companies should maintain an inventory of computerized systems, showing those that are GxP regulated." [2] Further guidance per *ISPE GAMP 5 (Second Edition)* [2] applies; see Section 6.2.8 in this ISPE AI Guide.

6.2.6 Planning for Validation

"Computerized system validation within a business unit is typically performed using a hierarchical framework of plans [(the validation plans)] covering GxP regulated computerized systems." [2] Guidance in *ISPE GAMP 5 (Second Edition)* [2] applies; AI-specific aspects are discussed in Section 9.3 in this ISPE AI Guide.

6.2.7 Continual Improvement Activities

"Improving the processes used to achieve and maintain compliance and fitness for intended use and making them more effective and efficient is highly desirable. The risk of non-compliance is also reduced." [2] When using AI, particularly ML, the potential for continual improvement of the process by using new data should be evaluated as part of ongoing monitoring and periodic review activities.

See Appendix P3 and *ISPE GAMP 5 (Second Edition)* [2].

6.2.8 Maintaining a Data and Model Inventory

Similar to system inventories (see Section 6.2.5), regulated companies should maintain data and model inventories, including documentation on the mapping between:

- AI-enabled computerized systems and their embedded models
- Models and data used to create, evaluate, or test these models

Information on the relevance and use of data and model inventories is provided in Appendices M7 and M8.

6.2.9 Responsible Use of AI Policies

Regulated companies should establish policies that promote the responsible use of AI in adherence to the trustworthy AI principles of Human Autonomy and Control, Safety and Security, Fairness and Mitigation of Bias, Privacy and Data Protection, Transparency, Accountability, and Sustainability.

Such policies should describe processes, responsibilities, and contributions of stakeholders integrated into life cycle activities and supporting processes; see Appendix M9.

6.2.10 Building AI Literacy

AI literacy involves a foundational understanding of how AI methods function, the considerations required to activate such innovation effectively, and their implications within the context of use.

Of note, AI literacy is a requirement per the EU AI Act [24]: *"Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used."*

Regulated companies should establish sufficient AI literacy among all stakeholders involved; see Appendix M5.

6.2.11 Data Ownership and Data Use Policies

Regulated companies should define data ownership in the context of AI-enabled computerized systems. Additionally, there should be alignment with other parties, including their suppliers, on what data can be used for what purpose; see Chapter 7. In establishing data ownership and data use policies, regulated companies should consider regulatory requirements and the specific process in which the data will be generated, and where it is intended to apply.

Further considerations on use of data are included in Appendices P1, P2, P3, P4, and M7.

6.2.12 Data and Model Governance

Data and model governance discussed in *ISPE GAMP 5 (Second Edition*) [2] applies to GxP data, or a data integrity perspective.

This Guide uses the terminology in a wider sense to describe not only technical aspects of data and model management and their governance, but also organizational aspects, such as roles or processes attached to those technical artifacts.

In this context, regulated companies should establish data and model governance and management as the organizational structures, policies, standards, and procedures concerned with data and models; see Appendix M7.

6.3 System-Specific Activities

Regulated companies need to perform various system-specific activities in the context of AI-enabled computerized systems. While many considerations concerned with non-AI-enabled computerized systems apply, specific aspects to these activities are outlined in the following sections.

6.3.1 Identify Compliance Standards

Regulated companies need to consider applicable company policies and procedures as well as national and international standards for non-AI-enabled computerized systems. [2]

6.3.2 Identify System

> *"The system should be assessed to determine whether it is GxP regulated and added to the system inventory in accordance with documented procedures."* [2]

6.3.3 Identify Key Individuals

While *ISPE GAMP 5 (Second Edition)* [2] provides general guidance on traditional roles such as process owner, system owner, data owner, quality unit, and end user, additional responsibilities arise in the context of AI-enabled computerized systems. There are also new roles relevant in the context of AI. These added responsibilities and new roles are discussed in this section.

While the exact definition of roles varies across organizations, the joint capabilities reflected by those roles are typically required to achieve safe and effective AI-enabled computerized systems.

Note: Not all roles described here may be required and some individuals may serve more than one role depending on the size of the organization and the project setting. However, holding multiple roles should not compromise required segregation of duties. The quality unit, for example, should always be independent from other roles on the project team.

6.3.3.1 Senior Management, General Management, and Leadership

The involvement of senior management, general management, and leadership roles vary throughout the life cycle, depending on the organization's structure and the significance of the use case. Typical responsibilities include:

- Understanding the potential of AI and anticipated benefits and risks
- Supporting prioritization and facilitating agreement about the planned AI projects
- Evaluating benefits, cost-effectiveness, feasibility, and risk tolerance
- Making go/no-go decisions based on information and insights collected throughout the AI project
- Facilitating internal and external dissemination of knowledge and learnings
- Fostering a fertile environment serving business objectives while maintaining quality

6.3.3.2 Process Owner

Process owners are responsible for the process the computerized system aims to support, managing functional requirements following the implementation roadmap. Responsibilities specific to the use of AI include:

- Understanding possibilities, limitations, and the potential impact of AI in their specific process to support decision-making for effective use of AI
- Navigating various options given the many possibilities of AI to support prioritization of feasible, high-impact use cases
- Integrating the implications of new data being captured on the system life cycle to achieve continual improvement and managing risks
- Deciding on options to iteratively improve the design of AI-enabled computerized systems, e.g., by use of further data sources or new models, or changing AI maturity (see Appendix M10)
- Ensuring sufficient control of the data in the operation environment; process owners are usually responsible for the data used and generated, and in some cases, become the de facto data owner (see *ISPE GAMP 5 (Second Edition)* Section 2.2 [2])
- Understanding the implications of choices of data and potential shortcomings on the process; their involvement in determining the fitness for purpose of data in the context of use is recommended; see Appendix M6

6.3.3.3 System Owner

"The system owner is responsible for the availability, and support and maintenance of a system, and for the security of the data residing on that system." [2] In the context of AI-enabled computerized systems, responsibilities include:

- Assessing high-level requirements for the infrastructure and performance including functionality with various volumes or loads on the infrastructure, such as model engineering, the generation of model output, and XAI methods intended to explain them
- Ensuring change management is performed as planned, deploying new models or managing the dynamic evolution of models in a dynamic system design
- Ensuring control of interfaces for supplying data serving as model input

To perform these tasks, system owners should possess a thorough knowledge of the applicable field, including its processes, challenges, business objectives, data needs, expected outcomes, and possible limitations of AI.

6.3.3.4 AI Sub-System Owner

AI sub-system owners are responsible for life cycle activities of the AI sub-system within the AI-enabled computerized system, including:

- Ensuring that the development of the AI sub-system and respective models serve the defined purpose within the context of use
- Ensuring integration of the AI sub-system into the AI-enabled computerized system, including use of model input and model output according to data specifications
- Overseeing the AI sub-system in operation, including ongoing monitoring and change management processes, for example, deployment of new model versions
- Aligning with other owner roles to synchronize and manage the AI sub-system
- Ensuring alignment with general organizational aspects pertaining to the AI sub-system
- Implementing measures to prevent misuse

6.3.3.5 Data Owner

Data owners are *"ultimately responsible for the integrity and compliance of specific data at various stages of the data life cycle in accordance with applicable policies and SOPs"* [15].

In the context of AI-enabled computerized systems and their AI sub-systems, data owners:

- Provide information to assess the representativeness, possible bias, and overall fitness for purpose of data, considering the context of use
- Establish an understanding of data generated during operation and its use for monitoring or continual improvement
- Provide insights into potential shortcomings of data, or changes in data conventions

While this role is not significantly different from non-AI-enabled computerized systems, it is a very important role, as the data owner collaborates closely with data scientists and data engineers to integrate process and data understanding in the design and development process of AI sub-systems.

The data owner may also be the process owner.

6.3.3.6 Project Manager

Project managers serve as the link between sponsors from senior management, project team members, and other stakeholders, and are responsible for:

- Managing innovation aspects of projects involving AI, often requiring an agile and iterative approach
- Facilitating openness and effective use of insights generated during the project and learnings from other projects
- Establishing realistic expectations on anticipated benefits of the AI approach sought in the project, and the value it will provide for the organization
- Balancing stakeholders' interests and objectives to facilitate comprehensive decision-making with respect to potential competing characteristics and objectives in the use of AI
- Managing the scope covered by the AI approach within the context of use, being mindful that a larger scope typically requires more data and activities to demonstrate the fitness for intended use

6.3.3.7 Business Analyst

Business analysts are responsible for communicating with business stakeholders and translating business requirements to technical team members in collaboration with the process owner and domain experts. They are responsible for:

- Eliciting requirements from stakeholders
- Defining the problem statement
- Determining the context of use
- Supporting the identification of suitable data sources and contextualizing results of the data assessment from the business process perspective
- Identifying suitable approaches, including those that use AI and those that do not
- Ensuring consistent understanding of the business process and the role of an AI sub-system and its models throughout the project

6.3.3.8 Data Scientist

Data scientists aim to provide value to the organization by managing, analyzing, and interpreting complex data. To improve processes and workflows, and to support data-driven decision-making, data scientists should be in close dialogue with data engineers, process owners, and data owners.

They hold the following responsibilities:

- Use statistical analysis, data cleansing, data mining, data modeling techniques, and AI to derive insights from data

- Develop data understanding and identify potential shortcomings in data, in collaboration with domain experts
- Establish prototypes to demonstrate the feasibility of an AI approach
- Support the AI project through iterative experimentation
- Provide expert support during operation, e.g., for RCA or deeper evaluation of data

A data scientist is not only a technical role; the profile typically requires an understanding of the business process as well as GxP and other regulations and requirements, such as data protection and cybersecurity, collaborating with SMEs and specialists in these areas.

6.3.3.9 Data Engineer

Data engineers prepare data for analytical or operational purposes, particularly for AI-enabled computerized systems. They hold the following responsibilities:

- Collect, store, and retrieve data in an efficient and scalable way
- Transform and curate data, considering the defined context of use
- Support data scientists in understanding the data
- Implement functionality and integrate technologies to support accessibility of data; this can include managing data storage implementations (data lakes and warehouses, etc.); see Appendix M8
- Implement and use tools for data analysis

Close cooperation with data scientists, process owners, and data owners is important in this role.

6.3.3.10 Data Steward

Data stewards hold an overarching responsibility for data, its governance, and for coordinating other roles related to data, for example, data owners, data scientists, and data engineers:

- Ensure the correctness and accuracy of data under their stewardship
- Help define and enforce data-related operational rulesets, including policies, SOPs, and guidelines; management and oversight activities go beyond tactical coordination and implementation aspects
- Support achieving data that is fit for purpose in the defined context of use
- Support strategic data management and data governance, beyond data usage, management, and security; see *ISPE GAMP Guide: Records and Data Integrity* [15]
- Provide a data integrity focused perspective to support the selection, curation, and management of data
- Collaborate with the AI project and operational teams and personnel generating, managing, and handling the data [15]; see Appendix M7

6.3.3.11 ML Architect

ML architects translate business requirements into system architecture and designs. Responsibilities include:

- Ensure adherence to corporate guidance, leveraging available infrastructure
- Develop proposals for extending the architecture with the appropriate rationale; for instance, some model types are only available on specific infrastructure or via dedicated platforms
- Derive options for supplied models (e.g., foundation models) and interfaces to respective systems, as well as data architecture, that support the development and system life cycle
- Support other technical roles and business analysts in adding detail to the technical system design

6.3.3.12 ML Engineer and AI Engineer

ML engineers and AI engineers combine knowledge of requirements and technical expertise in AI to design, develop, and manage AI sub-systems and their integration into AI systems. The role of an ML engineer is more classically focused on training or fine-tuning of ML models, while AI engineers cover a more general scope of implementing AI solutions. Generally, they have the following responsibilities:

- Establish robust practices in model management that ensure traceability of model iterations
- Implement deployment procedures for integration of AI sub-systems into the AI-enabled computerized system
- Serve as support personnel for ongoing monitoring during operation, including correction of errors resulting from incident and problem management activities
- Implement changes with the goal of continual improvement through use of new data

6.3.3.13 Software Engineer

Software engineers help design, implement, and verify the non-AI functions of the AI-enabled computerized system. They are responsible for integration of the AI sub-system and its models into the AI-enabled computerized system in collaboration with AI-specific technical roles, for example, AI and ML engineers.

6.3.3.14 Tester

Testers design, specify, and execute test cases based on a risk management activity following a risk-based approach per organizational guidelines and procedures. They hold the following responsibilities in an AI context specifically:

- Evaluating the interaction between humans and AI, including the effectiveness of XAI methods
- Ensuring reliable integration into the AI-enabled computerized system and robust process execution
- Testing configuration and change management procedures, for instance, the retraining or deployment of new models within the AI-enabled computerized system
- Performing boundary testing, to ensure correct implementation of controls
- Recording and evaluating test results with appropriate rigor based on risks; see Appendix M3

6.3.3.15 IT Infrastructure and IT Services Units

IT infrastructure and IT services units support fulfilling IT requirements in the context of AI-enabled computerized systems. Responsibilities include:

- Support fulfillment of infrastructure-related non-functional requirements throughout the life cycle, including the adequacy of outsourced services, computing resources, security, and scalability.
- Enable the use of specialized infrastructure, Graphics Processing Units (GPUs) for example, as applicable for the context of use and the chosen AI approach.
- Managing varying requirements throughout life cycle phases in close coordination and oversight by the system owner. For instance: while a prototype could be created with limited IT resources, training a neural network can require enormous computing resources. The model may require only minimal equipment during operation, while XAI methods demand higher computing capacities, depending on the use case.
- Support the implementation of functionality to manage and transfer data for operational purposes.

6.3.3.16 Cybersecurity Specialist

Cybersecurity specialists can help ensure secure and safe data and system infrastructures:

- Identify potential adversarial attacks, with a particular focus on data and models
- Support deriving strategies for the detection of potential adversarial attacks
- Foster a streamlined and harmonized approach to fulfill cybersecurity requirements according to the organization's security posture and procedures
- Raise awareness of the sensitivity of corporate or personal data that may be involved in a use case

Further details on security of AI-enabled computerized systems are provided in Appendix S5.

6.3.3.17 Training and Learning Management Departments

Training and learning management departments need to consider specific aspects in the use of AI in their training programs. These include:

- Establish AI literacy among all relevant stakeholders; see Appendix M5
- Raise awareness on the limitations of models to foster an environment of critical oversight, including adequate use of XAI methods, if applicable
- Train users to provide effective feedback, and raise the relevance of such interactions for ongoing monitoring and further development activities
- Consider real-world scenarios to demonstrate capabilities and limitations in the use of AI
- Establish programs to upskill existing roles and develop new roles as described throughout this section

A realistic understanding of the capabilities and level of knowledge of the people who will be trained is important for successful training. See Section 6.2.3.

6.3.3.18 Ethics, Legal, and Data Protection Personnel

Ethics, legal, and data protection personnel support in addressing various concerns in the context of AI-enabled computerized systems:

- Usage of data, whether for model engineering purposes or in collaboration with suppliers; see Section 6.2.11 and Chapter 7
- Support regarding data privacy considerations or business-sensitive data
- Support in evaluating trustworthy AI principles in the context of use of the model, and in determining implications for the development; see Appendix M9
- Support in ensuring adherence to applicable laws and regulations; see Appendix S7

These functions closely interact with other business and technical roles.

6.3.3.19 Procurement Teams

Procurement teams are responsible for the legal and contractual relationship with suppliers. Responsibilities include:

- Identification of legal and contractual risks early in the relationship between a supplier and a regulated company; this includes the alignment on the use of data; see Section 6.2.11 and Chapter 7
- Include AI-specific quality management aspects when engaging with suppliers
- Facilitate collaboration with other stakeholders to perform further evaluation and assessment of the supplier, such as its maturity and the suitability of the approach offered; typically, this process is led by quality functions in collaboration with process and domain experts, data scientists, and legal departments; see also Appendix M2

6.3.3.20 Quality Unit

Quality unit functions ensure that all decisions are made based on established processes and comprehensive rationales, i.e., in compliance with current regulations and fully aligned with company policies and SOPs aimed at high quality. This emphasizes the importance of maintaining adequate records of all development steps. They are also involved in the conduct of internal audits.

Specifically, IT Quality requires AI literacy to perform their oversight role, see Appendix M5.

Quality representatives support the AI-enabled computerized system throughout the life cycle:

- Support to determine adequate rigor of concept activities.
- Support and approval of initial risk assessments, determining the system impact.
- Support planning activities aiming for regulatory compliance and adherence to the regulated company's quality standards.
- Providing insights and facilitating good practices throughout development in the project phase.
- Assess suppliers regarding data management and model development practices in addition to traditional audit aspects; this is typically performed by IT Quality representatives cooperating with other Quality functions; see Appendix M2.

- Involvement in the assessment of fitness for purpose, and the AI-enabled computerized system's determination of being validated for intended use. In particular, the consistency of acceptance criteria and testing results and outcomes, as well as any deviations from planned activities, should be evaluated to support decision-making activities including those of model requirements specifications and model testing results.

- Interact with regulatory agencies to demonstrate compliance with the AI-enabled computerized system. This is performed by IT Quality representatives in collaboration with other Quality functions.

Formal activities and involvement should be based on the level of risk; see Chapter 5.

6.3.3.21 End User

End users perform tasks beyond merely using the AI-enabled computerized systems to actively contribute to its quality outcomes and further development, such as:

- Support the design of the system by providing and contextualizing requirements and evaluating suggested XAI methods if applicable; see Appendix S4

- When systems are of a suggestive nature: maintain awareness that model output is meant as a suggestion, taking action to achieve a high-quality result where relevant; see Appendix M10

- If designed accordingly, the end user should give suitable, high-quality feedback to the AI-enabled computerized system.

End users should have a sufficient level of AI literacy, and an understanding of how the AI-enabled computerized system functions as well on its limitations.[11]

6.3.4 Requirements Specification

General considerations per *ISPE GAMP 5 (Second Edition)* [2] apply in that requirements specifications are the responsibility of the regulated companies, with the possible support of a third party or supplier. Requirements include model requirements specifications that allow a regulated company to determine the fitness for purpose in the context of use of AI sub-system and their models; see Appendix P2.

6.3.5 Determine Strategy for Achieving Compliance and Fitness for Intended Use

Regulated companies should perform an initial risk assessment that determines whether the system is GxP regulated, the impact of the system, while it informs further planning of additional risk assessment activities. Guidance is provided in Chapter 5 and Appendix M3.

System components should be assessed, including their GAMP software categorizations. Specifically for AI sub-systems, the model category and the AI sub-system's maturity level should be considered; see Appendices M3, M10, and M11.

"The regulated company should formally assess each supplier to establish their quality capability;" further considerations per *ISPE GAMP 5 (Second Edition)* [2] apply, with AI-specific considerations provided in Appendix M2 in this ISPE AI Guide.

Scaling activities to the level of risk is a key consideration when determining the strategy for achieving compliance and fitness for intended use; see Appendices M3 and M4.

[11] Example – End Users in Clinical Trials: Clinical trials are settings where a large heterogeneity in end users is expected. For instance, patients may be the direct end users, where data science expertise should not be expected, generally. Further information on the use of AI in the clinical trial setting is in the *ISPE eClinical Good Practice Guide* [58].

6.3.6 Planning

"Planning is an essential activity for any system development and should address all aspects, including activities that demonstrate compliance and fitness for intended use." [2] General considerations apply per *ISPE GAMP 5 (Second Edition),* with specific aspects provided in Appendix P2.

6.3.7 System Specifications

Various specifications may apply, including functional specifications, configuration specifications, and design specifications [2].

Model design specifications and model selection may also apply. Details are provided in Appendix P2, which should be read in conjunction with *ISPE GAMP 5 (Second Edition)* Appendices D1, D3, and D6 [2].

As for non-AI-enabled computerized systems, the extent of specifications should be based on the nature and complexity of the system [2], and risks.

Practical examples are discussed in Chapter 4.

In addition, regulated companies should perform design reviews to evaluate deliverables against standards and requirements, identify issues, and, where applicable, propose required corrective actions; see *ISPE GAMP 5 (Second Edition)* Appendix M5.

6.3.8 Development and Review of Software for Custom Applications

Compared to non-AI-enabled computerized systems, development processes concerned with models in the context of AI typically follow practices of iterative experimentation.

While typically relying on agile practices, iterative experimentation does not imply practices to iteratively develop software; see Section 2.4.4.

In addition, the integration of the AI-enabled computerized system requires further development activities, where guidance of *ISPE GAMP 5 (Second Edition)* can be used; examples are user interfaces, technical integration of models created during iterative experimentation, and functionality used for data management.

Code reviews may be considered to support the quality of technical artifacts, for example, software code or (model) configurations. Per *ISPE GAMP 5 (Second Edition), "The need for, and extent of, reviews of new software during development should be based on risk, complexity, and novelty. Such reviews should be performed by an appropriate SME, typically from the organization developing the software. The regulated company should ensure that corrective actions resulting from such reviews are tracked to satisfactory completion."* [2] The same applies to code reviews created to establish models, including their pre- and post-processing functionality.

Code reviews by the regulated company are not required for standard and configurable software products, while such reviews may be included in defined standards for custom development. [2]

Further information on development processes is located in Appendix P2. Practices may consider prototype implementation, to base decisions on success or failure of a PoC on a robust foundation; see also Appendix P1.

6.3.9 Test Strategy and Testing

"The regulated company is responsible for ensuring that the test strategy will demonstrate compliance and fitness for intended use." [2] Specific aspects of testing apply when using AI, such as model testing and specific testing activities performed for the AI-enabled computerized system as described in Appendix P2. The extent and depth of testing activities should be based on risk, complexity, and novelty; see Chapter 5 and Appendix M2.

6.3.10 Reporting and Release

> *"At the conclusion of the project, a computerized system validation report should be produced summarizing the activities performed, any deviations from the plan, any outstanding and corrective actions, and providing a statement of fitness for intended use of the system. …In some cases, specific computerized system validation reports may not be required."* [2]

Further guidance on reporting and release is contained in Appendix P2, in conjunction with guidance provided in *ISPE GAMP 5 (Second Edition)* [2].

6.3.11 Maintaining System Compliance During Operation

"The regulated company is responsible for maintaining system compliance during operation" [2], see Appendix P3.

6.3.12 System Retirement

The regulated company is responsible for maintaining system compliance when retiring a system; see Appendix P4 and *ISPE GAMP 5 (Second Edition)* [2].

7 Supplier Activities

7.1 Introduction

"Although the responsibility for compliance with GxP regulations lies with the regulated company, the supplier may have considerable involvement in the process." [2] This pertains to computerized systems in general, including AI-enabled computerized systems. This chapter builds on the guidance provided in *ISPE GAMP 5 (Second Edition)* [2] with topics specific to the use of AI to help suppliers meet the regulated company's requirements and expectations.

7.2 Supplier Products, Applications, and Services

> *"Suppliers provide a range of products, applications, and services for hardware, software, and related technologies including the provision of cloud-computing services."* [2]

Typically, AI-enabled software products exhibit a higher level of complexity, with possibilities to adjust configuration on models and models themselves to companies' needs. Even when considering a categorization of a software product (such as a chatbot with limited configurability as a standard product (GAMP Category 3)), it should not be assumed that software, or integrated models in this category are simple. Therefore, the software category should only be one factor to scale life cycle activities, see Appendix M3.

Software categorization according to GAMP should be seen as a continuum rather than implying strict boundaries. For instance, adjusting a prompt, i.e., the instructions, of an LLM within an otherwise configurable software product may exhibit a substantial change in model performance, hence could be seen as holding elements of a custom application.

Custom applications (GAMP Category 5) may also integrate external AI components such as integration with LLM.

Note: An AI-enabled computerized system may include AI sub-systems associated with multiple categories.

Suppliers should consider that regulated companies need to assess whether the use of AI and the chosen approach is suitable for their context of use. For this reason, third parties, such as implementation partners, should have information that supports the regulated company's assessment in addition to the education, training, and experience needed.

Three foundational elements supporting such assessments are AI literacy, product and process understanding, and data understanding.

Data providers may play a more important role in the context of AI-enabled computerized systems compared to those not using AI, providing data for model development activities or testing purposes, such as testing limitations of models or products.

Suppliers can expect high focus of regulated companies regarding their data management practices, including practices of their sub-suppliers, as those practices are highly relevant for the safety and effectiveness of their AI-enabled computerized systems.

7.3 General Supplier Good Practices

General expectations on supplier good practices apply per *ISPE GAMP 5 (Second Edition)* [2]. These are briefly described here alongside additional aspects and considerations relevant for AI-enabled computerized systems.

7.3.1 QMS

Per *ISPE GAMP 5 (Second Edition)* [2], the supplier should provide a documented set of procedures and standards, ensuring activities are performed by suitably competent and trained staff, provide evidence of conformance with defined procedures and standards, and enable and promote continual improvement.

The procedures and standards for AI-enabled computerized systems should include:

- AI-specific aspects of roles and responsibilities, including augmented typical roles (e.g., cybersecurity specialists) and new roles (e.g., data scientists)
- Data and model governance, including selection, curation, and use of data as well as management and use of models
- Design of AI solutions, including processes to establish a rationale for the use of AI
- Selection of model designs, including processes to manage newly available models
- Evaluation of models, including decisions on the use of performance indicators and information capture on model performance
- Change management, including model versions iterations and communication of the impacts of changes
- Standards and processes to handle AI-specific incidents and problems

General principles apply for the traceability and control of data, models and software, development change control, configuration management, training, documentation and information management and backups; see *ISPE GAMP 5 (Second Edition)*. This ISPE AI Guide does not prescribe a specific process; guidance in Appendices P1 and P2 may serve as an orientation.

Suppliers may consider AI-specific standards such as ISO/IEC 42001 [5] to augment their QMS in addition to infrastructure and IT service delivery approaches (e.g., ITIL [12]) and continuous improvement of development processes (e.g., CMMI [10]).

7.3.2 Requirements

"The supplier should ensure that clear requirements are defined or provided by the regulated company." [2] Additional aspects include consistency with the chosen development method, development, review and approval processes, change control, testability from the point of view of the regulated company, and the use of tools, see *ISPE GAMP 5 (Second Edition)* [2].

Specific to the use of AI, the regulated companies should provide:

- Requirements on the model performance, by use of suitable performance indicators for the regulated company's context of use
- Requirements on the scope of applicable models; for instance, certain model types may be excluded by company policies

- Requirements on human oversight and involvement, including measures to support human decision-making, for example, the use of XAI methods
- Requirements for the adaptiveness of the system, specifically whether a static or dynamic system is foreseen
- Non-functional requirements, including the expected scale of data and use of models
- Requirements on AI-specific cybersecurity considerations

Appendix P2 provides further details on typical requirements.

7.3.3 Quality Planning

> *"The supplier should define how the QMS will be implemented for a particular product, application, or service. This should include defining the life cycle model being followed and the project organization, activities, procedures, deliverables, and responsibilities for establishing fitness for intended use of the system. The approach may include prototyping or other software development techniques. The role of supplier [Quality Assurance] QA should be clearly defined."* [2]

As part of quality planning, prototyping is mentioned as valuable approach in *ISPE GAMP 5 (Second Edition)* [2]. Specific to the context of AI-enabled computerized systems, suppliers may take advantage of guidance provided in Appendix P1 on prototyping activities and outcomes that support the regulated company's decision to continue the implementation. Of relevance for successful prototyping and meaningful results is an understanding of data to be expected under real-world circumstances in the regulated company's context of use.

7.3.4 Sub-Supplier Assessments

> *"Suppliers should formally assess their sub-suppliers as part of quality planning. They also should be periodically reassessed in accordance with the QMS. The decision whether to perform an audit of their sub-suppliers should be documented and based on a risk assessment."* [2]

Suppliers should consider the nature of their sub-suppliers when performing sub-supplier assessments. These include:

- Provision of data, including data used for model development purposes or used for model evaluation
- Provision of models, including those provided as open-source or as integrated products
- Provision of infrastructure capabilities
- Provision of services

Transparency in the use of such suppliers is expected, so that regulated companies can base their decisions on their understanding of the risk level involved in alignment with their policies.

They may adopt similar considerations as provided in Appendix M2 for their sub-supplier assessments.

7.3.5 Specifications

> *"For product development, the supplier should document the functionality and design of the system to meet the defined requirements. This should cover software, hardware, and configuration."* [2]

The supplier should provide details on relevant data and models used as part of their design specifications. The chosen model category should be documented, including its implication on change management aspects; for example, a fine-tuning approach with mixed data sources typically leads to more complex change management procedures.

While a complete description of the mechanics of complex models may not always be feasible, a rationale for the chosen data and models should be included considering the purpose and the role of the models and their limitations.

Suppliers may take advantage of the guidance provided in Appendix P2 for communication of such design aspects, including the use of data cards and model cards.

Objective testing should be based on predetermined performance indicators and thresholds of model performance, allowing for a verification of fulfillment of regulated companies' model requirements specifications in their context of use; see Appendix P2.

7.3.6 Design Reviews

> *"Design reviews evaluate deliverables against standards and requirements, identify issues, and propose required corrective actions."* [2]

When using AI, design aspects such as choices of models and data should be reviewed and evaluated during the development process and on a regular basis. Newly available models, or more available data, may motivate changes in the selection of suitable models, thus initiating the change management process in alignment with the regulated companies.

The risk, complexity, and novelty of the AI solution should be considered when planning the depth and cycle of design reviews; see guidance provided in Appendices P1, P2, and P3 and *ISPE GAMP 5 (Second Edition)* [2].

7.3.7 Software Production/Configuration

> *"Software should be developed in accordance with defined standards, including the use of code review processes. Configuration should follow any defined rules or recommendations and should be documented."* [2]

In addition to software and configurations, suppliers need to thoroughly manage models and data in the context of AI. MLOps infrastructure can help to support processes, such as:

- Storage, collection, and processing of data
- Data versioning and management of model artifacts
- Tracking of iterative experimentation, including capturing information on design aspects and model performance
- Performing change management including establishing new model versions

Suppliers may take advantage of the guidance provided on typical infrastructure elements in Appendix M8, and data and model governance and management practices in Appendix M7.

Review of design decisions concerned with the use of data and models, and how insights generated during iterative experimentation were leveraged for further iterations, are considered good practices to derive high-quality models, considering the regulated company's context of use.

7.3.8 Testing

> *"The supplier should test the system in accordance with approved test plans and test specifications. The test specifications, when executed, should demonstrate that all requirements, functionality, and design have been met."* [2]

In addition to typical testing activities of non-AI elements of the product (e.g., module (unit) testing, integration testing, system testing), suppliers and regulated companies should plan for model testing as part of their testing activities.

Model testing aims to confirm that the AI sub-system, including its model and pre- and post-processing functionality, meets requirements, as demonstrated on data that is fit for purpose for the regulated company's context of use. A dedicated test data set should be established independent from data utilized during model development. Establishing such a test set may require collaboration between the regulated company and the supplier to meet expectations that it is fit for purpose in providing realistic performance measures for the real-world environment in which the system should be used. As such, further data sources or the use of synthetic data may be considered to construct a suitable test set, depending on the level of risk, complexity, and novelty of the AI approach.

An additional aspect is testing the human-AI interaction, such as the use of XAI methods. Suppliers should provide information on their testing activities on such elements, which may include performance benchmarks of XAI methods, or manual testing performed by the supplier.

Testing may also involve AI-specific cybersecurity aspects. Suppliers should describe their approach to determine limitations and vulnerabilities to AI-specific attack vectors, and provide the effectiveness of controls as evidenced by their testing activities. These may include the use of synthetic data to demonstrate robustness and stability of the model or system.

See Appendices P2 and S5.

7.3.9 Commercial Release

> *"System release to customers [('commercial release')] should be performed in accordance with a formal process that describes the criteria for release, responsibilities, records to be retained, and items to be released, including software, hardware and documentation. …Release notes defining fixes, changes, known problems, and new features should accompany each release, including minor releases and patches."* [2]

In the context of AI, criteria for release should include the fulfillment of model requirements specifications, that is, in how far performance expectations are met. Further information includes:

- Employed models and model types, or changes to these for new releases
- Considerations on the use of data for establishing models, such as the use of fine-tuning and use of data
- Performance indicators and information on the limitation of models

> *"[A] commercial release by a supplier is not a release into the GxP environment, which is a regulated company activity."* [2]

7.3.10 User Documentation and Training

> *"The supplier should provide adequate system management documentation, operational documentation, and training for both maintenance and operation in accordance with agreed contracts."* [2]

Specific elements of user documentation and training include:

- Interactions with model output, including the use of XAI methods
- Expectations of users when reviewing or correcting model output, if applicable
- Expectations of users when providing feedback, if appliable
- Functionality for ongoing performance monitoring
- When using dynamic systems: information on expected and unexpected behavior when models change throughout operation

7.3.11 System Support and Maintenance During Operation

"The supplier should support and maintain the system in accordance with agreed contracts." [2] These include operational change control, configuration management, patch management, incident management, documentation management, backup and restore, business continuity, disaster recovery, managing software product releases, training supplier staff, system maintenance, and security management, see *ISPE GAMP 5 (Second Edition)* [2].

Of relevance in the context of AI-enabled computerized systems are:

- Ongoing performance monitoring, including performance indicators and drifts or trending over time
- Incident and problem management, including specific incident scenarios to the use of AI, and their implications on further development and monitoring activities
- Operational change and configuration management, including model updates and notification of, as well as alignment with, regulated companies on planned changes
- Supporting periodic review activities, including suitability of models and changes over longer periods of time
- Support regarding AI-specific cybersecurity threats; see Appendix S5

7.3.12 System Replacement and Retirement

> *"The supplier should manage the replacement or withdrawal of products or service[s] in accordance with a documented process and plans. Sufficient notice of the retirement of a system or version should be given to regulated companies to allow them to plan for their required activities."* [2]

In alignment with regulated companies, suppliers should maintain traceability of model input, models, and model output in the retirement phase to support regulated companies in *ex post* assessment activities as needed; see Appendix P4.

7.4 Data Usage

Suppliers should establish clear contractual stipulations for their use of data. These agreements consider and balance safeguarding the data of regulated companies while allowing suppliers to enhance their data sets, models, or software products. Careful alignment with regulated companies is required. Typical permissions suppliers may hold include:

- **Aggregated Data Analysis:** Suppliers may use aggregated, anonymized data to conduct research or improve system functionalities. Typically, this use does not disclose data of respective users while it utilizes a large volume of data to aggregate insights via trends, anomalies, etc. Suppliers should demonstrate transparency on the analytics that they perform on such data; opt-out mechanisms may be required.
- **Benchmarking and Comparative Analysis:** Data may be used to perform benchmarking services, where end user's data (which includes data from the regulated company) is compared against an aggregate of anonymized data from similar users to provide comparative insights and best practices. Suppliers should align with the regulated companies on the use of data for this purpose, and how far such (anonymized) benchmarks can be shared.
- **Improvement:** Data may help to improve their models or software products and enhance the data provided. This includes establishing new model versions by use of real-world data. When doing so, suppliers should securely segregate regulated companies' data from the real-world data obtained from other sources or use of their product.

- **Model and Product Development:** With appropriate rights and protections, suppliers can use insights derived from end user's data to develop new products or refine existing models or software products.

Suppliers may expect audits conducted by regulated companies that aim to ensure data practices are adhered to, while ongoing monitoring should ensure that data meets quality expectations (see Appendix M6).

Regulated companies are typically responsible for the consequences of data breaches impacting sensitive data managed by the supplier. Therefore, suppliers are typically contractually obligated to inform regulated companies about data breaches.

Data usage is linked to data ownership defined by the regulated company; see Chapter 6.

7.5 AI-Specific Security Measures

Suppliers should establish robust security measures. These include controls to prevent cyberattacks specific to the use of AI:

- **AI-Specific Threat Modeling:** Suppliers should consider specific threats to AI-enabled software such as adversarial attacks, model theft, or data poisoning. Suppliers should perform threat modeling specific to AI to identify and mitigate these risks effectively.

- **Robustness and Security Testing:** Before deployment, models and software products should be tested to ensure they are robust against attacks and perform reliably under various conditions. This includes testing for vulnerabilities to adversarial inputs.

These measures should be seen in combination with other cybersecurity practices:

- Secure development practices

- Third-party risk management and supply chain security

- Data management practices

- Implementing access controls

- Establishing incident response and recovery plans

- Establishing information security management systems, including controls for establishing management review

- Considering standards and third-party certifications such as ISO 27001 [59], ISO 27002 [60], ISO/IEC 27017 [61], ISO/IEC 27018 [62], ISO/IEC 27701 [63], ISO/IEC 27799 [64], ISA/IEC 62443 [65], AICPA SOC 1/SOC 2/ SOC 3 [66], The NIST Cybersecurity Framework [67]

- Demonstrating sufficient transparency

Further information is found in Appendix S5.

8 Appendix P1 – Concept Phase

8.1 Introduction

This appendix describes the activities performed during the concept phase of the AI-enabled computerized system introduced in Chapters 3 and 4.

This appendix covers aspects of early identification of business needs or opportunities briefly, irrespective of whether these lead to the use of AI or not. It then focuses on aspects relevant for the concept phase once the use of AI is foreseen, leading to the conceptualization of an AI-enabled computerized system or potential enhancements to existing systems.

ISPE GAMP 5 (Second Edition) [2] states: *"detailed activities in this phase will depend on company approaches to initiating and justifying project commencement. Gaining management commitment to provide appropriate resources is an important pre-project activity."*

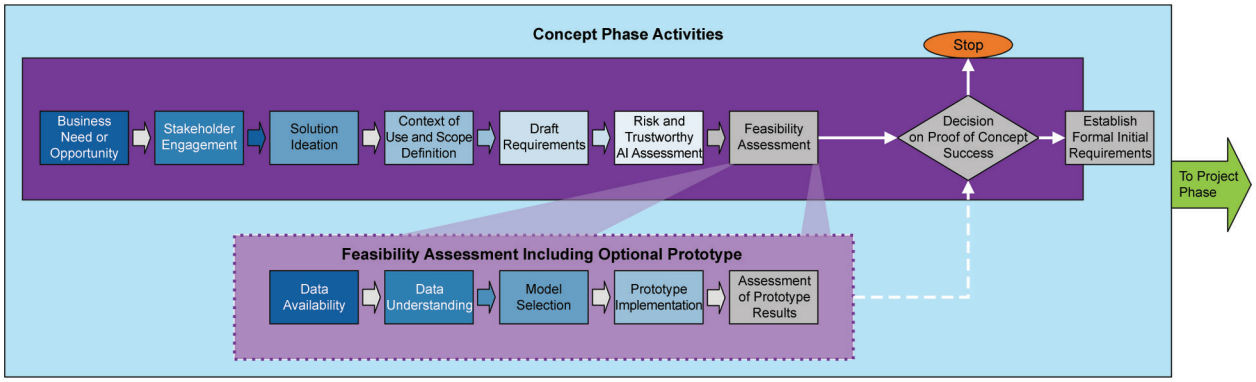
8.2 Overview

The concept phase initiates the AI-enabled computerized system life cycle and serves as the foundation for all subsequent activities. Activities include:

- Identification and assessment of the business need or opportunity
- Management of stakeholder engagement and planning of communication
- Ideation and selection of solutions
- Definition of the context of use and the scope covered by the solution
- Establishing draft requirements
- Performing initial risk assessment activities and assurance of considerations on trustworthy AI
- Conducting a feasibility assessment, which may include prototype implementation, leveraging model development practices
- Determining success or failure of the concept phase

Provided with a positive outcome, the capture of initial requirements completes the concept phase and allows for the transition to the project phase; see Figure 8.1.

Figure 8.1: Concept Phase Activities Overview



Note: This Guide provides suggestions on a typical sequence for these activities to allow for structured capture of artifacts and to facilitate success in subsequent activities. Many activities may be executed in parallel. There also may be situations where a prototype is developed in a less formal setting (e.g., in an internal challenge competition).

The concept phase aims for the following goals:

- Ensure alignment and effective contributions of stakeholders
- Identify and prioritize use cases that provide value to the organization; this includes achieving business goals and raising quality standards
- Determine a suitable approach to address a challenge or business need; this may or may not involve the use of AI
- Develop a rationale that underpins the expected potential of the chosen approach
- Identify suitable data sources and model development options to inform subsequent development steps
- Identify possible shortcomings in data quality
- Raise confidence in elevated efforts required in subsequent steps via demonstration of the potential of the chosen approach
- Identify relevant regulatory requirements for subsequent planning
- Prepare effective activities in the project phase via well-understood requirements

Along the life cycle of an AI-enabled computerized system, three entry points are relevant to enter the concept phase (see Figure 8.2):

- **Initial Entry:** This entry-point represents the situation when the system is initially designed or when a new AI sub-system should be added to an existing computerized system.

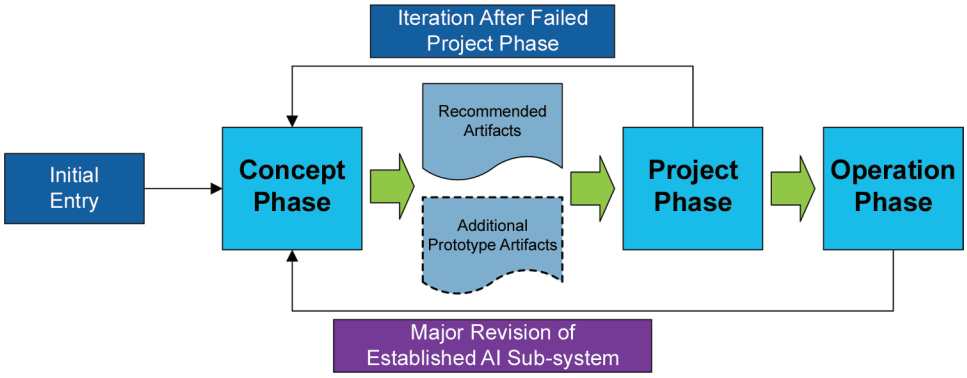
  In this scenario, SME input is important compared to other scenarios, as quantitative evidence on the system's use is limited or non-existent. Regulated companies are recommended to implement a prototype to generate first insights into the potential performance, limitations, and effectiveness of control strategies.

- **Major Revision of an AI-enabled Computerized System:** Like other systems, AI-enabled computerized systems may undergo a major revision from time to time, affecting their AI sub-systems. Examples for drivers that motivate major revisions are:
  - Availability of a new model type
  - Availability of more data
  - Change in AI maturity levels (e.g., reducing verification of every model output to reviewed by exception, see Appendix M10)
  - Need for revision because of lack of performance as observed during ongoing monitoring in operation (refer to Appendix P3)

  Regulated companies typically focus on adapting existing components. Good practices also involve a review of the overall process, risks, and the control strategy, leveraging experience from operations.

- **Iteration of the Concept Phase after a Failed Project Phase:** Upon failure in the project phase (e.g., when expectations on performance are not met), regulated companies should perform a careful evaluation of whether to conduct another iteration of the concept phase instead of declaring the project failed. They should review historical decisions as a basis for an improved design and redirect the project towards success.

Figure 8.2: Concept Phase Entry Points



8.3 PoC Structure

Concept phase activities should be organized by capturing information in a structured way, aiming for deliverables that support decision-making throughout the phase and eventually determining its failure or success.

Good practices include stepwise capture and refinement of information, collectively becoming a PoC, while deciding on key activities as more is learned about the problem, business need, and use case. For instance, a crucial decision to be made is whether to conduct prototype implementation.

The typical information captured in the PoC is outlined in Table 8.1. The choice of applicable elements depends on the organization's experience using AI and the novelty of the approach, while hurdles identified through PoC activities may suggest additional activities and assessments.

Table 8.1 Key Elements of a Typical Proof of Concept (PoC)

Key Element	Details on Records and Information
If Applicable	
Business Need/Problem Statement	The business problem to be addressed
Planning for Stakeholder Communication	Strategic communication plan to reach alignment and inform relevant stakeholders about intended changes
Stakeholder Engagement	Details stakeholders for the project and their contributions
Use Case Definition	Goal to be reached
Solutions	Possible approaches that fulfill the use case definition; upon selection of use of AI, the Use Case is seen as an AI Use Case
Context of Use and Scope Definition	Definition of the context of use and scope
Initial Draft Requirements	Define the intended use of the system and its key requirements
Trustworthy AI and Alignment with Human Values	Assessment of adherence or challenges regarding trustworthy AI principles
Initial Risk Management Activities	System impact, initial capture of hazards, and ideas for controls; evaluation of business risks
Feasibility Assessment	• Whether a system offers a benefit that is in good economic proportion to the costs • Initial assessment of availability of suitable data and identification data shortcomings
PoC Acceptance Criteria	Acceptance criteria that determine success or failure of the PoC
PoC Summary	Summary of whether the system passed or failed to meet the acceptance criteria
If Prototyping is Performed	
Data Availability Assessment	How the assessment of availability of suitable data for the system is performed
Data Quality Assessment	Suitability of data to meet key quality requirements
Initial Considerations on Model Selection and Suitability	Selection of feasible model options, covering an indication of developing new models versus using or fine-tuning pre-trained models
Prototype Results	Technical aspects of prototypes and insights into their performance

Records and information on PoC activities are not part of GxP validation records but become an input to them. For instance, they inform the capture of initial requirements.

The concept phase relies on the diverse expertise of various stakeholders.

Suppliers may also serve different roles in supporting organizations in the context of AI-enabled computerized systems. Regulated companies may take advantage of:

- Insights about previous successful AI projects
- Demonstrating AI capabilities in the use case setting

- Moderating discussions among stakeholders to bridge technical and business domains
- Executing assessment activities or establishing prototypes
- Assembling information used for management decision-making
- Identifying synergies and dependencies with other AI projects, and supporting roadmap planning
- A collaborative co-development setting to extend existing software products

8.4 Business Need or Opportunity

In the initial step, regulated companies identify the problem or opportunity that should be addressed. It involves activities such as

- Problem framing
- Alignment with business objectives
- Requirement gathering
- User and stakeholder interviews
- Solution and supplier exploration
- Scope definition

Activities aim for incremental refinement, deriving a Business Problem Statement and an AI Use Case Definition.

8.4.1 Business Problem Statement

The business problem statement defines the challenge in the process that should be addressed by the use of technology. Its articulation is crucial for its downstream implications in the concept phase and later life cycle phases.

The statement should outline the problem concisely and describe the specific issues and concerns intended to be solved. The problem statement should not include or explicitly propose or suggest a specific solution. Product quality or patient safety aspects should inform the problem statement, including determining potential high-level risks, and consider any specific GxP regulatory expectations or guidance, see Appendix S7.

Stakeholder involvement is relevant to elicit more detailed needs and determine a relevant problem statement that is well understood. The problem statement should be refined as needed as more understanding is gained in subsequent steps.

A poorly defined problem statement, and/or poor communication and coordination among stakeholders may jeopardize the success and/or adoption of AI by business units, and lead to possible business and quality-related risks.

Good practices include capturing ideas worthy of consideration for inclusion while prioritizing business problem statements that carry high relevance to stakeholders.

8.4.2 Use Case Definition

Once clarity on the business problem statement is defined, business objectives are identified and should be met. Metrics should be selected intended to determine if business objectives are fulfilled. Examples include saving time, reduction of errors, and reduced carbon footprint.

A use case should help articulate Specific, Measurable, Achievable, Relevant, and Time-bound (SMART) objectives, including to what extent an opportunity should be seized, or a challenge should be addressed. These are still agnostic to the solution approach. For instance:

- "Machine X is rejecting more vials during automated inspection since DD MMM YYYY than previously reported by a factor of X. We need to determine the root cause(s) of the elevated levels of vial rejection to help ensure we can supply our patients without adding unnecessary cost to the product via false rejections."

- "Manual (human) tagging of data during input has been demonstrated to be incorrect up to 50% of the time. We need an automated method for determining data integrity related deviations from a population size of several thousand records per year that does not rely on human manual data tagging and provides an accuracy of at least 80% to help target internal audits, without having to use manual text searches."

Alignment of use cases with business objectives is important to focus efforts on projects with the most significant impact. This in turn helps to ensure resource optimization, such that time, funding, and human resources are used efficiently to avoid wasted or redundant effort on use cases not aligned with business objectives.

Prioritizing use cases along a strategic direction provides these benefits:

- Management of dependencies between use cases
- Achievement of synergies
- Promotion of stakeholder buy-in
- Ensuring communication and alignment
- Facilitating collaboration

Aligning use cases with business objectives enables a line of sight between the risk of a given AI-enabled computerized system and the potential business benefit. When evaluating benefits and risks, the organization should take a life cycle perspective, considering efforts throughout all phases.

The outcomes of use case definition activities include:

- The business objective that is supported by the use case, which informs the purpose of the intended component
- Outline of the expected business benefits
- Understanding of dependence to other use cases or projects

8.5 Stakeholder Engagement and Planning of Communication

Sufficient stakeholder engagement during the concept phase allows for the contribution of diverse views. The engagement of stakeholders varies throughout the concept phase.

Additionally, a communication plan, which helps coordinate the stakeholders, should describe the feedback processes and stage gates to facilitate stakeholder alignment.

Good practice involves establishing a safe space in which concerns can be voiced, providing the project team with constructive feedback from various perspectives, enabling them to address issues as they arise. Stakeholder engagement and planning of communication adjust throughout the concept phase as more insights are generated on the relevance and possible contributions of stakeholders.

8.6 Solution Ideation

Once the use case definition is agreed to, stakeholders typically ideate possible solutions. Factors relevant to meaningful ideation are:

- Experience of the team
- Technological possibilities
- Maturity of the organization
- Established systems
- Available data
- Input from suppliers

Solutions are identified and assessed. Upon consideration of AI methods as part of the solution, the Use Case converts to an AI Use Case, which becomes the primary focus for subsequent activities.

Examples of use cases that may warrant AI-based automation include:

- Repetitive and time-consuming tasks that require an intelligence component
- Tasks that occur frequently and involve manual effort
- Tasks that require complex decision-making based on copious amounts of data
- Tasks prone to human error that can be completed more reliably and consistently with an AI approach, e.g., image processing
- Use cases that can be scaled and used across different scenarios or business processes
- Risks identified through the incident management process where traditional approaches offer limited possibilities to mitigate risks further

8.7 Context of Use and Scope Definition

The goal of the context of use is to determine the role AI, and subsequently models, serves in the process. Scoping determines which areas within the context of use are expected to be supported by the solution. Key determinants of the scope include:

- Business problem statement and its relevance in various areas of the use case
- Availability of data and quality of data that represents these areas
- Experience of users
- Experience of the organization
- Risks

- External requirements, such as regulations
- Suppliers' capabilities and available functionality in their software products

These considerations may lead to exclusions and limitations in the context of use. These should be captured to avoid use in unintended areas that may exhibit higher risks and where risk mitigation strategies may not be adequate.

The context of use and the scope typically are refined during subsequent activities.

8.8 Drafting Requirements

The initial requirements are drafted to make the requirements generally understandable to all stakeholders. The initial collection does not need to be captured in a formal record. They are meant to facilitate discussion and guide subsequent steps in the concept phase (see Section 8.12 for formal capture of requirements).

This involves defining both functional and non-functional requirements, and should address aspects such as scalability, security, performance, and usability to generate high-level requirements. A simple example is described in Table 8.2.

Table 8.2: Example Problem Statement and Draft Requirements

Problem Statement	"Manual (human) tagging of data during input has been demonstrated to be incorrect up to 50% of the time. We need an automated method for determining data integrity related deviations from a population size of several thousand records per year that does not rely on human manual data tagging and provides an accuracy of at least 80% to help target internal audits, without having to use manual text searches."
Draft Requirements	1. The solution should run over data from the Deviation Management IT system stored in the organization's data lake. Functionality to load data should be verified, and the data storage should be qualified. 2. The AI sub-system should classify data to identify data integrity related deviations in a binary manner (data integrity related – Yes or No) using an ML model, giving a probability score in %, with an accuracy of 80% or over. 3. The AI sub-system should identify subtopics within the deviations such as data falsification, missing data, non-contemporaneous data recording, data lacking attributability, and data accuracy issues. 4. The AI-enabled computerized system should display the probability score against each deviation record and group the data Topic Modeling output to enable identification of recurring themes. 5. The use of this AI sub-system should speed up the searching and data listing process by X% and clustering process by Y%.

Initial consideration should be given to the following non-functional requirements and system design aspects, with further details typically refined and formally determined in the project phase:

- **System Performance**
  - Time to load and allow display of data
  - Cycle in which data is refreshed in the data storage
  - Computational resources required

- **Security and Privacy**
  - End-to-end security of the data flows and systems
  - Access control on data and data representations in interfaces
  - Operational controls
- **Usability**
  - Ease of use of interfaces
  - Adequacy of information provided to end users
- **Reusability**
  - Reuse of models or data
- **Choice of Static and Dynamic Systems Designs**

8.9 Initial Risk Assessment and Trustworthy AI Considerations

Managing, identifying, and assessing risks begins early in the development process. This appendix is concerned with risk management in the concept phase. Further information can be found in Appendices P2 and M3.

Trustworthy AI principles should be considered; see Appendix M9. In addition, risks should be assessed by dividing them into two categories: Quality and business-related risks. Figure 8.3 illustrates the scope of recommended assessments.

Figure 8.3: Initial Risk Management Activities and Trustworthy AI Considerations

8.9.1 Consideration of Trustworthy AI Principles

Alignment with human values is assessed by means of trustworthy AI principles. By integrating trustworthy AI principles early in the concept phase, activities in subsequent phases can consider those human values and mitigate potential risks to meet expectations of external stakeholders.

Stakeholders should evaluate whether the AI use case's objectives and potential outcomes are consistent with fundamental ethical principles and societal norms. Their impacts on human value and rights should be understood beyond potential patient safety and regulatory impacts.

The following list of questions illustrates relevant aspects that regulated companies are recommended to cover in their evaluation during concept phase:

- **Human autonomy and control:** Does the level of control and insights on model input and output foreseen for end users allow them to sufficiently perform oversight in the context of use?
- **Safety and security:** What safety implications does the AI sub-system have on individuals? What implications are expected from compromises of the system?
- **Fairness and mitigation of bias:** Does the use case exhibit unfair treatment or discrimination of certain sub populations?
- **Privacy and data protection:** What restrictions on the use of data need to be met for subsequent life cycle phases?
- **Transparency:** How should activities in subsequent phases be executed to ensure sufficient understanding of the AI-enabled computerized system for all relevant stakeholders?
- **Accountability:** How can responsibilities be organized throughout subsequent phases that maintain and promote awareness of accountability?
- **Sustainability:** What implications and possible side effects may impact society disadvantageously?

8.9.2 Initial Risk Assessment

Relevant aspects of the initial risk management activities include:

- Performing the initial risk assessment and determining the system impact
- Gathering information on functions with impact on patient safety, product quality, or data integrity
- Capture of initial considerations of possible hazards and ideas for controls

Experience and insights from past– successful or failed – PoC activities and AI projects should be leveraged, while ensuring diverse engagement of stakeholders.

Chapter 5 and Appendix M3 contain additional information.

8.9.3 Business Risks

Business risks should be evaluated to enable robust cost-benefit analysis and informed decision-making on the most promising use cases. These include:

- **Investment risks:** AI projects may require considerable upfront investment and resources. There is a risk that these investments will not yield the expected return. There is also the challenge of achieving a return on investment within a brief period, especially when an organization is entering uncharted territory with AI applications.

- **Organizational risks:** Organizational risks include inadequate management support for the development and deployment of AI sub-systems, and a lack of experts, such as data scientists or ML and AI engineers. The integration of AI systems into existing processes and systems can entail operational challenges, including the need to change workflows, train staff, and adapt to the IT infrastructure.

- **Project management risks:** These risks relate to the planning and execution of AI projects. In addition to insufficient resources, unrealistic schedules, and a lack of project planning and management, etc., there are risks specific to AI. These risks arise from the experimental nature in developing AI, managing unknowns regarding the interplay of data and models for the use case.

Go/no-go decision points are common, providing opportunities to decide whether a project should be continued or not. If it is likely that the necessary objectives and requirements of a project will not be achieved, a timely cancellation limits economic losses.

These decision points in the development process and corresponding decision criteria should be defined in the concept phase. Such go/no-go goals can stop a project in early phases. Regulated companies should capture reasons for the go or no-go decision to ensure complete tracking of the project and to facilitate dissemination of learnings.

Typical decision points throughout the ideation process include:

- Prioritizing relevant business problem statements in early ideation of use cases

- Prioritizing use cases that admit a clear rationale for the use of AI

- Prioritizing use cases that allow for clear measures of success aligned with the organization's objectives

- Prioritizing use cases with sufficient readiness of the organization and stakeholders

- Prioritizing use cases with satisfactory data availability and quality

- Prioritizing use cases that show satisfactory performance or exhibit only minor limitations in prototypes

- Prioritizing use cases that elicit positive feedback from end users when demonstrating prototypes

- Prioritizing use cases with clear responsibilities

8.10 Feasibility Assessment

Conducting a feasibility assessment analyzes the advantages and challenges of the AI use case in greater detail. A feasibility assessment assists in decision-making, supporting an understanding of:

- Benefits, providing more certainty and robustness of expectations
- Organizational preparedness including management support, acceptance, trust, training
- Stakeholder involvement and alignment of interests
- Context of use, including end users and their challenges or needs, infrastructure restrictions, and data availability
- Economic assessment of project-associated risks and their mitigation
- Required resources and access, e.g., appropriate personnel, IT capacities, computational resources, data sources
- Selection of suppliers and available products
- Regulatory understanding, assessing areas where regulatory advice is needed

Consideration of designs that are not based on AI may serve as the basis for providing a comprehensive rationale on the chosen approach.

8.10.1 Data Availability Assessment

Assessment activities should cover the availability of data relevant to the AI use case. These should identify potential shortcomings and understanding of the impact of such shortcomings on the feasibility of the solution.

The assessment of data availability addresses:

- Access to historical data that can be used to train or evaluate a model
- Identification of sources to provide suitable data
- Assurance of access to these data sources in the future
- Availability of interfaces to integrate data for prototyping and later phases

The effort in answering these questions to a sufficient extent depends on the organization's maturity in managing data (Appendices M6 and M7) and the defined scope within the context of use.

Planning further activities is based on the data availability assessment because if the result is a negative evaluation, the use case may not be realized using this data.

8.10.2 Development of Initial Data Understanding

Initial understanding of available data is developed by exploration and analysis. Typical practices include:

- Visualizations
- Statistical evaluations

- Use of AI

The aim is to identify correlations and relationships in the data that can be used to support the AI use case.

One method to increase data understanding is Exploratory Data Analysis (EDA). It involves the initial exploration of data sets to gain insights and identify patterns and relationships, spot anomalies, test hypotheses, and uncover potential issues.

EDA includes:

- **Understanding individual variables and metadata** by calculating frequency counts, visualizing distributions, and summarizing the data.
- **Examining relationships between variables and metadata** using visualizations, deriving correlation coefficients, and other statistical techniques.
- **Identifying trends and patterns** that facilitate understanding of the underlying data structure.
- **Detecting outliers and anomalies** that could indicate data issues.

EDA provides several advantages during the concept phase:

- Understanding the data structure, and becoming familiar with the data's format, variables, and the relationships between the variables, is central to building data understanding.
- Determining shortcomings of data quality, such as missing values, outliers, or inconsistent data types, and gaining initial insights on patterns, trends, and correlations to inform the choice of models and features.
- Data understanding in conjunction with process and product understanding enables hypotheses generation for further analysis preparing for later model development activities.

Visualizations from EDA results help communicate complex data insights to stakeholders and facilitate discussion to combine insights from data with domain expertise.

Gaining data understanding typically is an iterative process in which teams refine and adjust the results. Feedback from SMEs helps to integrate domain knowledge with data understanding.

Although such preliminary work is intended for data understanding and is not seen as a formal process, the steps to obtain the data set used as basis for the assessment should be captured. Steps include data collection, data evaluation and transformation methods, and analysis or visualization methods. Relevant metadata should be captured for traceability purposes, e.g., the data collection point in time. Good practices use version control for software code.

Techniques mentioned in Appendices M7 and P2 may be used for further data analysis.

8.10.3 Model Selection

Possible model choices are proposed in the concept phase, which will be refined in the project phase. Determinants of model choices include:

- The intended use and high-level requirements
- The context of use

- Data availability
- Data quality and suitability
- Available models
- Experience of the regulated company
- Information from suppliers

The chosen models may also be used for prototype implementation during the project phase.

8.10.4 Prototype Implementation

A key method for assessing the feasibility of the AI use case and assessing risks is to develop a prototype. A prototype also helps in requirements elicitation and identifying end-user needs; see ISO/IEC/IEEE 29148 [68].

A prototype is a simplified version of an AI sub-system, but explicitly developed, runnable in a way to indicate possible performance and demonstrate feasibility and thus, guide decision-making. It can also help to gain further data understanding and insights into possible shortcomings of data quality.

While the prototype typically would not suffice the rigor of a GxP environment, it informs the model design during the project phase.

When developing a prototype, similar steps are taken as for the AI sub-system, see Appendix P2; however, the level of formality is typically reduced, commensurate with business and financial risk.

Good practice includes the use of tools that allow flexible working, support the experimental nature of the activity, and provide a variety of functions that allow data manipulation and analysis, and can incorporate a wide range of models. Several tools are available, including open-source software or integrated into large development platforms from software suppliers. Depending on the use case, the following features can be helpful to achieve efficient prototyping, (see Appendix M8):

- Interactive use is facilitated, possibly leveraging low- or no-code functionalities
- Line of code can be executed interactively
- Results can be displayed in parallel
- Changes immediately lead to the most recent results
- A variety of graphical representations are possible
- A large selection of statistical functions and ML models is available

The right choice of tool should be determined based on development activities requirements, corporate standards, available data, and models considered for use.

Prototype implementation may also involve alternative designs, including designs not based on AI, as a basis for providing a comprehensive rationale for the eventually chosen approach.

Insights should be summarized and documented as they can support the more formal activities in the project phase.

8.11 PoC Evaluation Criteria

Agreed and communicated PoC evaluation criteria enable transparent and comprehensive decision-making to prioritize and decide among various choices.

Evaluation criteria to be fulfilled for the PoC constitutes a positive decision on progressing the use case and the basis for transition into the project phase. Common practices include using scorecards that assess insights gained from concept phase activities. Typical areas of evaluation include:

- Expected business value
- Readiness of the organization
- Evaluation of risks
- Data availability
- Data quality
- Infrastructure and technology readiness
- Performance demonstrated via prototypes
- Assessment of end-user feedback
- Supplier maturity, if applicable
- Synergetic effects on the digital and data landscape

Management typically receives evaluations and scores in structured reports as the basis for their decision-making, underpinned by relevant insights from the prototype, as applicable.

8.12 Specifying Initial Requirements and Transition to Project Phase

At the end of the concept phase, initial requirements are established as the starting point for more formal development in the project phase.

Requirements can be set up per *ISPE GAMP 5 (Second Edition)* Appendix D1 [2], informed by considerations outlined in this ISPE AI Guide to cover the particularities of AI sub-systems and their integration into the AI-enabled computerized system.

9 Appendix P2 – Project Phase

9.1 Introduction

This appendix describes the activities performed during the project phase of the AI-enabled computerized system introduced in Chapters 3 and 4.

Compared to the development of non-AI-enabled systems, the activities to include AI are predominantly iterative, providing stepwise improvement and advancing understanding of the interplay of data and models with each iteration. Iterative experimentation comprises model selection, data engineering, model engineering, and evaluation.

The outcome of a successful project phase is the successful release to operation of an AI-enabled computerized system holding one or many AI sub-systems, marking the transition to the operation phase.

ISPE GAMP 5 (Second Edition) [2] should be followed for developing integrating functionality, for example, interfaces to models. This appendix focuses on activities for model development. In combination, models and integrating functionality construct an AI sub-system. For the development of non-ML AI, this appendix refers to *ISPE GAMP 5 (Second Edition),* where applicable; the choice of an adequate approach depends on the purpose and complexity of such non-ML AI models.

This appendix primarily addresses regulated companies who seek to develop high-quality models, being aware of risks that may evolve from decisions during model development with potential impact during operation. For technical details regarding the implementation of models, activities of the project team are discussed. Other stakeholders (for example, suppliers) may take advantage of the guidance provided regarding the quality expectations of regulated companies.

Static and dynamic systems are in scope of this appendix, highlighting nuances in development approaches.

9.2 Overview

The project phase is characterized by planning, design and specification, iterative experimentation processes, and verification activities to eventually determine the fitness for intended use and release of the AI-enabled computerized system.

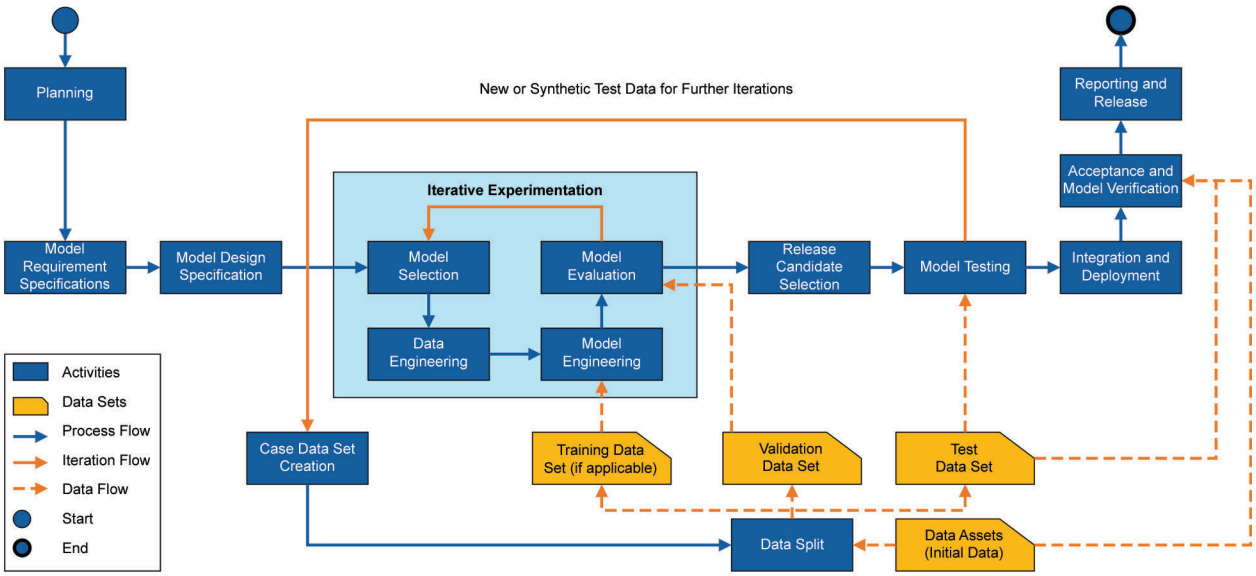
This phase allows creativity to explore ways to address business challenges, while maintaining awareness of the importance of achieving a high-quality AI-enabled computerized system. As shown in Figure 9.1, key activities include:

- Planning
- Determining Model Requirements Specifications
- Determining the Model Design Space
- Defining Key Metrics and Thresholds
- Creating the Case Data Set
- Performing Data Splits
- Conducting Iterative Experimentation
- Defining Model Release Candidates

- Performing Model Testing
- Executing Acceptance Testing
- Reporting and Release

Key goals are to delineate the technical architecture, manage risks and implement effective control strategies, and develop models and their embedded AI sub-systems that support the intended use of the AI-enabled computerized system.

Figure 9.1: Project Phase Activities Overview



Note: The guidance provided for development activities is not meant to be a prescriptive approach; instead, it serves as a baseline of common practices, while regulated companies and teams may choose other activities or sequences of activities.

9.3 Planning

Planning leverages results from the concept phase to develop a detailed understanding of activities needed in the project phase. It aims for scaling activities according to the level of risk and concludes by demonstrating the AI-enabled computerized system's fitness for intended use.

Important aspects of planning are:

- Roles and responsibilities, considering domain and technical expertise and sufficient AI literacy
- Risk assessment activities including functional risk assessments to identify AI-specific hazards and suitable controls
- IT infrastructure with particular focus on access to model development environments and models
- Planning data management activities, including the creation of case data sets and test data sets that serve model testing purposes

- Planning iterative experiments, including the rigor on capture of information and records of individual experiments
- Alignment with suppliers on supplied data, services, or AI-enabled software products
- Acceptance criteria specific to AI sub-systems
- Acceptance criteria for the functionality related to AI within the AI-enabled computerized system, e.g., human-AI interaction, and change management processes

Initial requirements and initial risk considerations established in the concept phase inform these activities.

Communication of the results of planning activities to all relevant stakeholders is important to perform subsequent activities efficiently and effectively.

9.4 Model Requirement Specification and Definition of Model Design Space

The model requirement specification and the definition of model design space aim to identify all requirements and outline model development options that correspond to the requirements and the outcomes of the risk assessment.

Functional and non-functional requirements are specified. These specifications serve as the guideline to ensure resulting models adhere to performance benchmarks and are fit for purpose. Then, a design space is determined that exhibits admissible options for developing models tailored to meet those requirements and mitigate potential risks.

The model requirement specification and definition of the model design space build on results from the concept phase, including:

- Business problem statement
- Use case definition
- Results from the data availability assessment
- Results from the data quality assessment
- Results from prototype implementation, if feasible
- Initial requirements, which include ethical, regulatory, or legal aspects

A key design aspect is the choice of adaptiveness, and in particular the choice between static and dynamic system design. If a dynamic system design is considered, requirements should be established that not only control the performance of the model but also control its evolutionary learning path.

Relevant information and records include:

- Model requirements specifications, in particular KPIs and metrics
- Description of the model design space including choices of third-party provided models
- Rationale of why the defined model requirements specifications and the design space are suitable given the context of use within the AI-enabled computerized system

These inform the iterative experimentation planning activities.

9.4.1 Model Requirements Specification

Regulated companies should establish a set of formal requirements that align with the designated objectives per the requirements captured from the concept phase. The AI sub-system requirement specification encompasses both functional and non-functional requirements.

- **Functional requirements** pertain to the specific operations and context of use, i.e., the role of employed AI sub-systems and their models in the process

- **Non-functional requirements** are the broader characteristics that include the operational context of a model. For model design and operations, typical non-functional requirements include fault tolerance, reliability, and efficiency. For model retraining and revision, they can include testability, re-trainability, and scalability, and for model transition, they can include complexity, reproducibility, and interoperability, see Habibullah, Gay, and Horkoff 2023 [69] for further details).

The standards for capturing such requirements are the same for functional and non-functional requirements.

In the case of a dynamic system design, further requirements arise regarding the adaptive nature of such systems (see Appendix M10). These include:

- Specifying controls on the evolutionary development path of models. These build on model requirements specifications, although they take a dynamic perspective. For instance, for an automated retraining being performed and set live, requirements may include a performance loss of no more than X% should be observed across a particular stratification (i.e., monthly slices of data in a manufacturing context, or types of patients in a diagnostic use case).

- When considering a range of potential competing models within a dynamic system context, clear rules describing under which circumstances what competitive model should be (automatically) selected to allow for traceability and comprehensibility of the model selection process.

- Additional requirements include data preprocessing, iterative or continual model learning, automated verification of model performance, rollback functionality, and infrastructure needs as well as further requirements specific to model types.[12]

9.4.1.1 Key Performance Indicators and Thresholds

The selection of KPIs used to measure performance characteristics relevant to context of use should be linked to product and process understanding, data understanding, and the results of risk management activities.

The selection of suitable KPIs relies on collaboration among multiple disciplines, including technical and domain expertise, while leveraging experience from past projects. While some performance indicators are commonly used, tailored or adjusted indicators may apply when justified by a rationale on why they capture specific performance characteristics relevant to the context of use.

Given the number of performance indicators available, and their impact on decision-making regarding the AI sub-system, a rationale on the choices of performance indicators is needed, captured as part of the project phases' information and records. The following illustrate typical considerations on the strengths and weaknesses of KPIs; they should not be understood as a complete list:

[12] Some models (like those generating text or images) offer the possibility to apply watermarks, i.e., features within the model output that are typically difficult to remove. This may constitute a requirement in cases where the verification of the content authenticity of content is important.

- **Classification:**

  - *"Accuracy is the proportion of all classifications that were correct, whether positive or negative."* [70] It provides a baseline measure of the overall model performance, although it is insensitive to a) class imbalances (i.e. whether one class within the set of classification is more prominent than others) and b) differing risks associated with differing errors.

  - Recall, or true positive rate, is the *"proportion of all actual positives that were classified correctly as positives."* [70] If risk is primarily associated with the ability to correctly identify positive cases (e.g., actual defects of a product), more weight may be given to this performance indicator.

  - *"Precision is the proportion of all the model's positive classification that are actually positive."* [70] This measure may be used to guide decisions on human verification for positive cases (e.g., to determine whether positive samples are indeed defective), in case precision is relatively low.

  - The F1 Score, i.e., the harmonic mean between precision and recall, provides a measure that dedicates equal relevance to precision and recall, irrespective of the proportion of the respective classes' observations [70]. It provides a metric measuring the overall performance of the model, adjusting for class imbalances.

- **Regression:**

  - The mean absolute error is the average absolute difference between predicted values and actual values, which requires a distance measure for the output of a model.

  - The mean squared error is the average of the differences between predicted and actual values. Mean squared error between predicted and actual values increases weighting of large deviations while decreasing small errors.

- **Image Segmentation:**

  - The Hausdorff distance is a measure used to compare the similarity of boundaries of a segment of an image or higher-dimensional structure. In its simple form, it measures the maximum distance between the respective closest points defining a boundary. [71]

  - The Dice score can also be used to measure the overall segmentation performance of a computer vision model, accounting for the typical lower share of the image or higher-dimensional structure of interest. It is defined as twice the number of common elements of the predicted and a reference segmentation mask, divided by the sum of the numbers of elements in the predicted and in the reference mask. [72]

  - The Hausdorff distance is concerned with boundaries; the Dice score focuses on the area correctly identified, augmenting each other in guiding decisions on models.

Other performance indicators include measures between model output and actual outcomes for modalities such as generated text (for instance, BLEU scores in the context of translation [73] or ROUGE score to compare summaries from an input text [74]), or for other applications (e.g., the Concordance Index may be used to measure the alignment of model's predictions with the occurrence of events [75], useful in the context of survival analysis for a clinical trial).

These considerations also include the definition of an end criterion for the incremental experimentation and iterative feedback process to allow for an objective decision and selection for a model release candidate. For example, it can be defined that the performance metric reaches a specified limit or that the metrics of a tuning sequence no longer change significantly.

These KPIs inform requirements on performance monitoring in the operation phase.

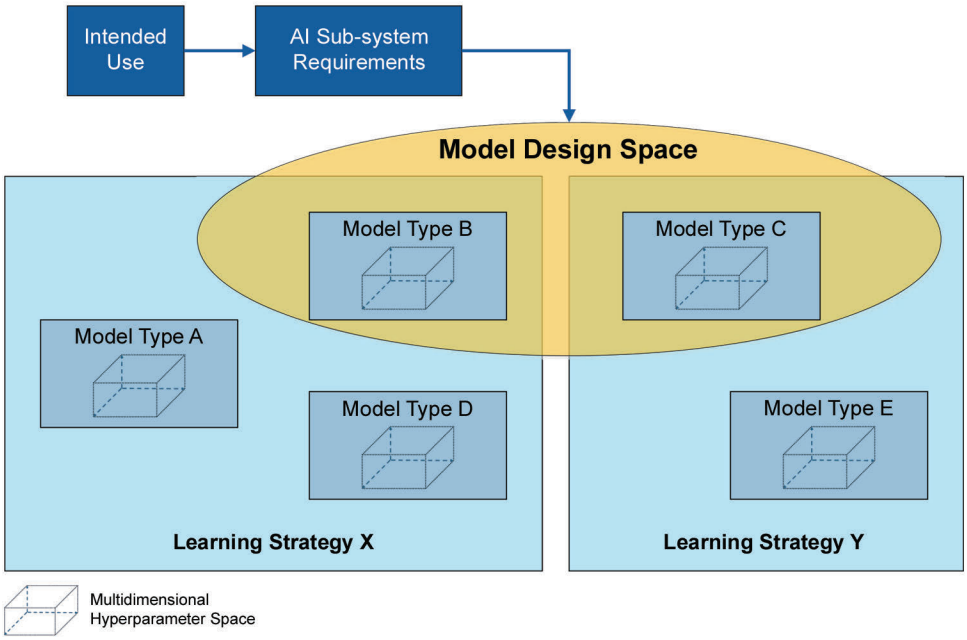
Leveraging metrics and thresholds, processes should be specified, covering aspects such as:

- Real-time monitoring or real-time detection
- Data input verification
- Data quality assessment
- Anomaly detection
- Bias monitoring
- Alert notification
- Human feedback

9.4.2 Model Design Specification

The model design space specifies the range of applicable learning methods; see Figure 9.2. The goal is to select a space within the wide range of model types or architectures under consideration for subsequent development steps. Defining the model design space to prevent missing high-potential model candidates needs to be balanced against the increased costs needed to cover a wider space. A selection of learning strategies and model types for consideration in the model design space is provided in Appendix S3.

Figure 9.2: Illustration of a Model Design Space



9.5 Case Data Set Creation

The case data set collects data obtained from data sources identified during the concept phase or during preceding activities in the project phase. Fitness for purpose supports the quality of models developed by the use of this data in subsequent activities.

The following sections present necessary steps to make a case data set technically available, to assess the data's fitness for purpose, to convert it into harmonized formats, and, if necessary.to enrich it with additional data.

Applying these steps considers the context of use, requirements, product and process understanding, model design and prototyping results, where applicable. Data is collected using qualified infrastructure and data pipelines that were determined to be fit for purpose.

9.5.1 Data Collection and Creation of the Case Data Set

A suitable case data set is created, ensuring the following is true:

- The case data set is fit for purpose in supporting the development and testing of models with respect to the context of use and the model design space.
- The processes to extract, curate, and augment existing data are captured, enabling reproducibility (e.g., by using identifiers, keys, or timestamps) and traceability of data.
- If data augmentation techniques are used, their fitness for purpose is ensured.

The fitness for purpose of the case data set depends on model requirements specifications and the model design space, since more complex models typically require more data for subsequent activities.

Data management practices as outlined in Appendix M7 support achieving a high-quality case data set that is fit for purpose for subsequent activities.

Data requirements for dynamic system designs are similar to those for static systems during the initial development process. In addition, functionality for the automated provision of suitable data ("data pipelines") is planned to support adaptive learning. Such data pipelines include the following steps:

1. Obtain data from defined sources
2. Perform data quality assurance
3. Apply transformations
4. Integrate with further model engineering and evaluation procedures

Quality gates are implemented throughout this process that trigger human interaction when data expectations are not met. Such quality gates leverage product and process understanding and may include:

- Checking for missing values
- Inconsistent data types
- Incorrect or implausible entries

See Appendices M6 and M7.

Beyond individual characteristics of data points, monitoring of shifts or drifts in data distributions is assessed.

9.5.2 Data Enrichment

In preparation of subsequent development steps, the obtained data set may be enriched with further information and characteristics. Feature engineering and data labeling may apply.

Feature engineering is the process of creating new features from existing data to enhance analysis or model development efforts. Feature engineering uses domain knowledge and applies analytical skills to extract and create new features (variables) from raw data that make models work more effectively. Feature engineering involves several techniques, including:

- **Transformation:** Applying algorithmic transformations to variables to change their distribution or relationship with the target variable, or to prepare for improved effectiveness in serving as model input.
- **Creation:** Deriving new features from existing data. For example, this can involve combining two or more data fields or extracting parts of data fields.
- **Selection:** Choosing the most relevant features to use in model development. This entails identifying and removing irrelevant, redundant, or highly correlated features that do not contribute to or may decrease the model's performance.

Data labeling provides labeled data for supervised or semi-supervised learning strategies; see Appendix S3. Since data may not be equipped with labels when originally collected, labeling or annotation augments the case data set. Options to provide labels include:

- Annotation by humans according to predefined standards
- Fusion of data sources to augment data with labels
- Use of models to establish labels or support humans in identifying the ground truth; the risks for using such techniques should be evaluated and their fitness for purpose ensured.

9.6 Data Split

Data split prepares a selection of data sets to perform a fair, unbiased evaluation of the performance of models in subsequent steps during iterative experimentation. In the following, two typical approaches for data splitting are discussed: simple splitting into training, validation, and test data sets and the more complex cross-validation technique. Both exclude test data from the data used for iterative experimentation to derive an estimate of how the model's performance will generalize on new, unseen data (see Section 9.9).

"Unseen" means:

- Unseen by the model in terms of training (if applicable)
- Unseen by the project team in terms of performance characteristics already evaluated

A test data set may also be expanded by additional independent data (e.g., from other sites) or synthetic data to detect possible limitations of models, depending on the risk (see also Section 9.9).

In a dynamic AI sub-system approach, data splitting needs to be fully automatized, including considerations on ensuring the fitness for purpose of the resulting data sets. The data splitting techniques described here for static systems can also be used for dynamic systems, while other suitable approaches may apply depending on the context of use.

9.6.1 Splitting into Training, Validation, and Test Data Sets

The case data set is split into a validation set and test data set, in addition to a training data set if applicable. The purpose of these sets is:

- To be used in an iterative experimentation and feedback loop (see Section 9.8), i.e., the two data sets are possibly used many times:
  - **Training data set** (if applicable): This data is used to train a model from scratch or to refine an existing model.
  - **Validation data set:** This data is used to evaluate the model's performance after a model engineering step to guide the further development process or allow for a determination to stop the iterative experimentation process.
- To be used ideally once for final verification of the model's performance
  - **Test data set:** This data is supplied to the set of model candidates for final assessment of performance characteristics via KPIs (see Section 9.9)

To reasonably approximate the performance that can be expected in the operation setting, the test set needs to be representative for the context of use of the model it intends to test (see Appendix M6). Representativeness also supports the development of a profound model via the training data set, and guides the decision-making process based on the iterative model evaluation using the validation data set. Techniques such as Analysis of Variance (ANOVA) tests and other distribution comparison methods can help demonstrate the representativeness of data.

Interrelationships between those data sets should be considered. One example is "spurious data leakage," i.e., situations in which the validation, and more importantly the test set, is not independent from the other data sets. For example, some observations may share common characteristics in a biopharmaceutical manufacturing case if they are sourced from a single seed. A second example of such leakage considers possible duplications across the validation, training, and test sets in a clinical trial setting where various data sources have been fused to a single case data set. The occurrence of data leakage should be evaluated regarding possible risks, based on product and process understanding, and data understanding.

When considering dynamic system design, the iterative use of training, validation, and test data sets is adapted to accommodate the ongoing collection and preparation of data. The purpose of these sets in a dynamic context is:

- **Training data set:** This set is used to iteratively train or fine-tune the model. It is periodically refreshed or enhanced with new data.
- **Validation data set:** This set is used to evaluate the model's performance after automated retraining or fine-tuning. Techniques such as rolling validation windows, pronounce more recent observations, if suitable in the context of use.
- **Test data set:** Ideally used only once for final performance verification when deriving a new model version, the test set is periodically refreshed. Failure to demonstrate performance on the test data set typically triggers a shift into the static mode of the system, including a notification for human interaction.

9.6.2 Cross-Validation (K-Fold Cross-Validation)

Cross-validation is a technique for evaluating models and testing their performance. K-fold cross-validation splits the data set into K equal-sized subsets, trains the model on K-1 subsets, and tests the remaining fold. This process is repeated K times, and the results are averaged to assess model performance.

The goal of a cross-validation approach is to provide wider insights into a model's expected performance on unseen data rather than on a single data set. This provides insights into the uncertainty or robustness of the KPIs, which can be informative regarding the model design (inner fold cross-validation) or the final test (outer fold cross-validation), contributing to the control strategy of change management activities during operation.

This approach may be favored when holding only limited data or when uncertainty is expected in KPIs, causing higher computational cost and time consumption as model engineering and evaluation runs multiple times.

9.7 Iterative Experimentation

The goal of this activity is to discover models that yield optimal performance for the context of use. This is achieved by engineering and evaluating multiple model instances compared to selected performance indicators and thresholds in the model design space.

Iterative experimentation is typically conducted in four steps, repeatedly executed (shown in Figure 9.3):

1. **Selection of suitable model design:** Selection of a model type, relevant data within the case data set, and a set of hyperparameters required to establish the model

2. **Data Engineering:** Preparation of suitable data that fits the selected model

3. **Model Engineering:** Technical implementation to establish the model

4. **Model Evaluation:** Derive KPIs and compare the new model against the existing set of previous models

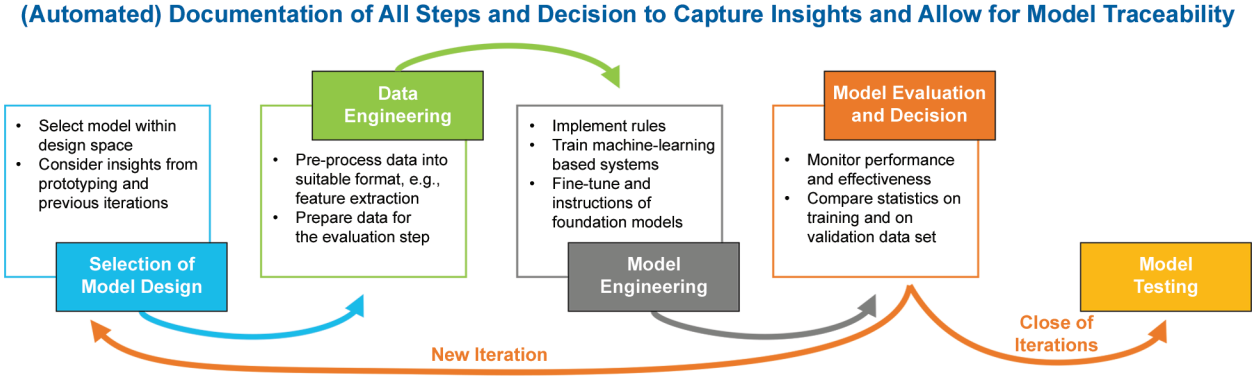
If the model does not achieve the desired thresholds for model performance indicators in the model evaluation step, the team conducts further iterations. They leverage understanding gained throughout the process to inform data engineering or model engineering in the next iteration.

The process is repeated until a consensus is reached on the modeling approach that produces the best-performing outcome, reaching a stop criterion. Information and records on these experiments, including the data used as model input, model selection, hyperparameters, configurations, and model performance indicators should be captured in a traceable way to allow *ex post* assessment.

Data and Model Cards

As part of the system specification, teams should capture key technical elements of data and models. One way of providing the relevant information on a high level is via data and model cards, summarizing key aspects necessary to understand the information the data represents, and the architecture and purpose of the model. When collected and managed in a streamlined way from the wider perspective of the organization, data and model cards can be used to a) select suitable data or models from previous activities, and b) provide new data and model cards as those artifacts are created during the incremental experimentation; see Appendix M7.

Figure 9.3: Iterative Experimentation Overview



In a dynamic AI sub-system context, iterative experimentation does not only cover the model designs but also control designs that govern model adaptiveness. These include aspects such as:

- How often automated data engineering and model engineering procedures should be performed to derive a new model version
- What performance triggers capture the dynamics of changing environmental conditions best
- What controls could safeguard the AI sub-system's evolution

Both aspects, the model choice, and the choice of the dynamics of model evolution and its controls, should not be seen as singular, separate activities, but rather evaluated in conjunction with respect to the model's context of use and the dynamics of real-world relationships expected during the operation phase.

The iterative experimentation process concludes with a set of models, their specifications, and their evaluation results that exploit the model design space according to the stop criterion. These form the basis for defining a release candidate model.

Note: Exploitation of the model design space does not mean that all possible configurations have been used to derive models, which is typically impossible given the degrees of freedom. For this reason, *ex-ante* defined stop criteria help to determine the conclusion of the iterative experimentation phase, while critical thinking should be applied based on the insights generated during iterative experimentation, process and product understanding, and data understanding; see Appendix M4.

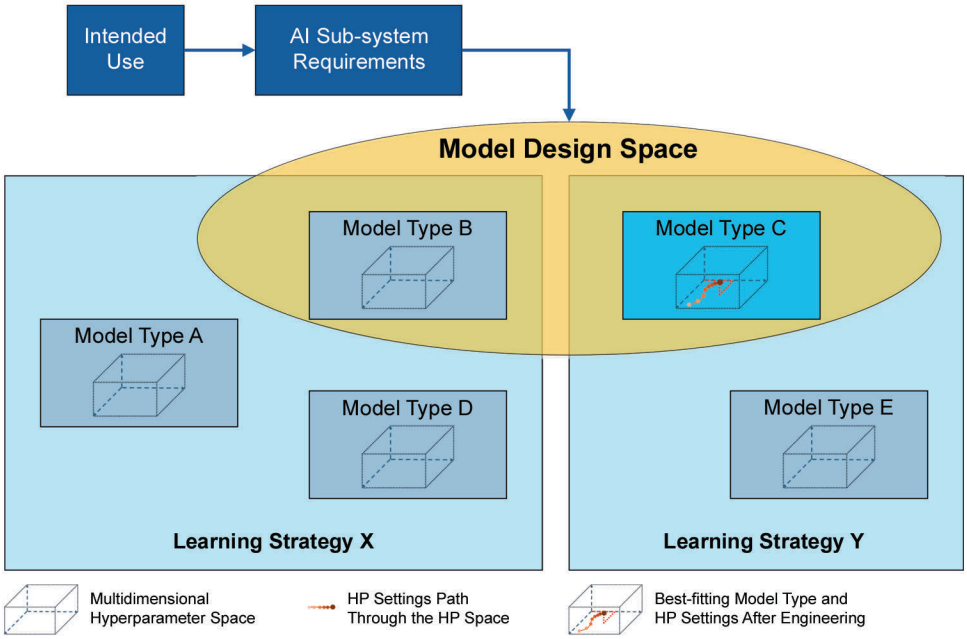
9.7.1 Model Design Selection

Teams should select a model design that aligns with the problem it aims to address, considering the model design space and model requirements specifications. Model selection is influenced by possible trade-offs between concurring objectives (e.g., precision and explainability) and should be based on the following:

- Previous iterations (if any)
- Prototype implementation in the concept phase (if applicable)
- Product and process understanding and data understanding generated throughout the project phase

A model design consists of applicable data and all relevant hyperparameters, as shown in Figure 9.4. Theoretical analysis or preliminary testing with data batches may support model design selection before entering further full-scale data engineering and model engineering as well as evaluation steps.

Figure 9.4: Illustration of a Best-Fitting Model Within the Model Design Space



An important part of the model design is the choice of hyperparameters. Hyperparameters are parameters of the model that define the architecture and control the training process of a model. The process of choosing the values of hyperparameters to optimize model performance is referred to as hyperparameter tuning. Each type of model has a set of hyperparameters that is relevant to its architecture. Each of these hyperparameters has its own definition range.

The range of values of all hyperparameters span a multidimensional space referred to as the Multidimensional Hyperparameter Space in Figure 9.4. By varying the values of its hyperparameters, the properties and performance of the model vary.

The goal of hyperparameter tuning is to determine the best set of hyperparameters for a particular model to maximize its performance given model requirements specifications.

Common problems in training models, including overfitting, can be avoided by choosing appropriate values for the hyperparameters of a model. Various hyperparameters can be applied to a model for performance improvement. For example, Lasso and Ridge regression are two techniques used to enhance the performance of regression models. Both methods involve adding a penalty to the objective function; the magnitude of this penalty is the hyperparameter that can be tuned. The hyperparameter controlling the strength of this penalty is applied to the model's coefficients. In turn, the hyperparameter reduces the model's probability of fitting to irrelevant noise in the training data, thus improving its generalization to new data.

Similar hyperparameters are also used in other ML techniques, such as the depth, split, and number of features in decision trees; the weights and number of neighbors in K-Nearest Neighbors (KNN) (see Appendix S3); and the learning rate, batch size, number of hidden layers, and dropout probability in neural networks.

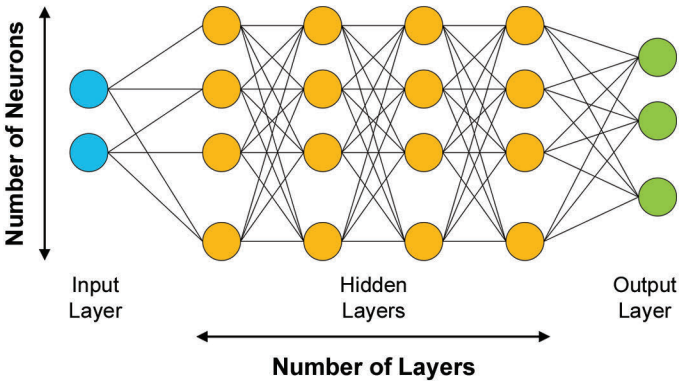
As an example, as illustrated in Figure 23.5, feed forward neural networks have several essential hyperparameters that need to be adjusted, including:

- **Number of layers (hidden layers) and neurons per layer:** Experiments with different architectures is a way to find the right balance between complexity and performance when choosing a higher number of layers.

- **Number of training epochs:** The number of times the entire training data set is passed through the model during one training cycle. An additional patience hyperparameter can be used to determine how many epochs to wait before stopping training once performance no longer improves.

- **Batch sizes (number of batches used for one adjustment of the net):** The number of training examples used in each iteration. Figure 9.5 shows the first two hyperparameters.

Figure 9.5: Basic Neural Network Architecture Example



9.7.2 Data Engineering

The objective of data engineering is to prepare data for model training and evaluation purposes. Model requirements specifications and the selected model inform data engineering, defining the format of data that is suitable as model input and, if relevant as is the case for supervised learning methods, the format of the labeled data.

The activities performed during data engineering are as follows:

- **Preprocessing** the data to a format suitable to be input to the model for training and evaluation. Typically, this will involve engineering the data so that it is more easily ingested by the models.

- **Post-processing** involves transformation of the model output to facilitate the calculation of qualitative performance metrics or to generate qualitative performance outputs.

9.7.3 Model Engineering

The goal of model engineering is to establish a model that captures the essence of real-world mechanics as a sufficiently accurate representation. Model engineering is based on the following considerations of preceding steps:

- Model requirements specifications and the model design space

- Available data sets pertaining to the chosen data split strategy

- Data prepared by data engineering steps

The following activities are typically considered for model engineering:

- Preprocessing of data, in addition to processing in the data engineering step

- If training or fine-tuning of models is involved, technically defining the model and its hyperparameters as per model design selection step; a model is derived by applying the training technique

- Defining the technical integration of data and models, e.g. selection of features or, particularly when employing foundation models like LLMs, implementation of the prompting strategy
- Possible further post-processing steps of model output, based on considerations in the data engineering step

Further considerations are provided for three forms of model engineering in the following sub-sections:

- Implementation of rule-based systems, considered as non-ML AI; this category is included for completeness
- Use of ML
- Use of foundation models

Note: Combinations of the above approaches are possible. For instance, a rule-based system operating on some features may be coupled with an ML model on a set of further features. Hence, these forms should not be seen in isolation, but as a flexible set of options to achieve a high-quality overall model.

9.7.3.1 Implementation of Rule-Based Systems

Implementation of rule-based systems includes manual implementation or configuration of rules. *ISPE GAMP 5 (Second Edition)* [2] provides guidance regarding development practices of such systems.

9.7.3.2 Use of ML

To train a model, three elements are decisive for a given model type: the model input as provided by the training data set, hyperparameters defining the model architecture and the learning behavior, and an objective function.

During model training, data is first ingested, to derive the model by learning patterns from this data. The objective function, often called a cost function, serves as the model's teacher. The objective function provides a measure to evaluate the performance of the model on a task (e.g., match between real-world ground truth and model output). Training aims to minimize the objective function by choosing suitable parameters. For example, the least squares method [76] minimizes the difference between the model output and actual values in the training data set.[13]

Various models exhibit the choice of batch training techniques. Such techniques update model parameters incrementally over multiple passes of data. Each batch is a subset of the training data set. As each batch of the training data set is introduced to the model, the model's parameters are changed incrementally.

Complex models, such as some neural network architectures, require these batch optimization techniques. In contrast, simpler models, such as linear regression, offer analytical methods to minimize the error of the objective function and hence directly retrieve the optimal set of parameters given the model.

After batch training, validation data and test data sets can be used per usual to assess the model's performance.

[13] An objective function can be considered appropriate if it reflects the balance between various performance considerations based on business and quality objectives. For example, high deviations between ground truth value and model predictions may constitute a higher risk, so that (if feasible) an objective function should be chosen that is super-linear with respect to those differences. If, however, the average difference should be minimized, a linear relationship should be considered.

9.7.3.3 Use of Foundation Models

In many cases, teams may choose pre-trained foundation models to avoid initial costs in developing a model. Teams may integrate such models with fine-tuning by using limited data to meet requirements or use foundation models without any further fine-tuning via integration (e.g., prompting techniques in the case of LLMs or following a RAG approach; see Appendix S3). Both approaches significantly reduce the training time and computational resources needed to derive a model.

While purely integration-focused approaches do not alter the parameters of the model, fine-tuning involves taking a pre-trained foundation model and adjusting it with a smaller, task-specific data set. This process customizes the model to perform well on a particular task or with data types to align closely with stakeholder processes and requirements.

While the use of foundation models can expedite the development process, rigorous testing remains crucial with or without fine-tuning. Pre-trained models may not always capture the nuances of a new application domain without adjustments. Teams should evaluate the necessity to perform fine-tuning with data specific to the use case, weighing it against its implications on the complexity of change management processes (see Appendix P3) in alignment with the regulated company.

9.7.4 Model Evaluation

Following model engineering, the model's performance is evaluated. Performance is measured using selected performance indicators, derived from model input and model output when applying the model on prepared data sets pertaining to the chosen data split strategy. Performance indicators are called in-sample when applied on the data set used for training or fine-tuning purposes, and out-of-sample otherwise.

While these indicators convey a model's effectiveness on a particular data set, differences in indicators when applied on training versus validation data sets provide insights into the model's generalizability. For example, a model may perform well on the training data set and achieve exceptional performance. However, when this model is applied to the validation data set, the performance may be poor. This could be an indication a model is overfit on its training data set, meaning it has learned the minutia of the training data set which is not representative of real-world data. Alternatively, a model may exhibit a poor performance already when applied to the training data set, with relatively small differences in performance when applied to the validation data set. This indicates an underfit model, meaning it may not be able to capture the real-world relationship between model input and model output, or requires further training data to improve its effectiveness.

The interpretation of model performance results should consider the risks and context of use. Visualizations such as charts or tables help in comparing model performance, e.g., for various subsets of data or across models derived throughout the iterative experimentation process. Statistical tests and confidence intervals may provide a quantitative evaluation of differences in model performance and its uncertainties. By observing the variability of a model's performance results, insights may be derived regarding the robustness and overall generalizability of the model.

Evaluating the model enhances product and process understanding and data understanding, providing insights into the interplay between model input, models, and their output. This enhanced understanding translates into ideas for further improvements, including suggestions for adopting new model architecture or performing further data or model engineering as input for the next iteration. For example, if a model's performance indicators exhibit a tendency for overfitting, a recommendation may be to increase the data set size and change hyperparameters to improve overall model generalization. In this regard, the EMA reflection paper [43] suggests techniques such as drop-out or regularization.

Additional measures like the training duration may support the planning of iterative experimentation or may be informative for future training and fine-tuning of models during the operation phase (see Appendix P3).

9.8 Model Release Candidate Selection

Based on experimentation results from iterative experimentation, a subset of models called model release candidates is selected for integration into the AI-enabled computerized system.

Models derived during iterative experimentation are excluded when their KPIs do not meet thresholds. Meanwhile, the most suitable models are selected by assessing and comparing their design, complexity, and performance metrics. The assessment and comparison of dynamic system designs and their resulting models should not only include the individual model performance but also the learning behavior.

Model version control as well as visualization techniques to compare performance across models support the model selection process. Records and information on the chosen models and their hyperparameters, as well as pre-processing and post-processing techniques, should be maintained. Records and information on decisions related to model release candidate selection should be captured, to allow for alignment and transparency in line with model requirements specifications and the defined design space.

9.9 Testing of AI-Enabled Computerized Systems

To achieve AI-enabled computerized systems that are safe, effective, and fit for intended use, testing strategies and the rigor to be applied need to be determined, based on the level of risk as per functional risk assessments and impact on the process and relying on critical thinking. Testing activities need to adhere to the regulated company's policies and procedures.

Appendix D5 in *ISPE GAMP 5 (Second Edition)* [2] describes general aspects for testing of computerized systems, providing information on roles and responsibilities, influences on testing activities, test planning and coverage, types of testing, test environments, leveraging previous testing including supplier testing, and testing applied to different software categories. In addition to these general considerations, the use of AI introduces additional aspects for AI-enabled computerized systems:

- Integration of the AI sub-system owner and other technical experts, such as AI and ML engineers, into testing activities, in collaboration with other roles
- Influencing factors and AI-specific nuances related to test activities include:
    - Functional risk assessments specific to the use of AI (see Appendix M3)
    - Design of the AI sub-system including its maturity with implications for change management processes, in particular functionality for model version updates
    - Availability and fitness for purpose of data (see Sections 9.5 and 9.6) in the context of use of the regulated company
    - Reliance on performance indicators derived from application of models on test data sets that are fit for purpose
    - Requirements that are specific to the use of AI such as human-AI interaction or AI-specific cybersecurity threads
    - Changing model output for the same model input as characteristic of some types of models, which needs to be considered in determining the objective when creating test cases

- AI-specific aspects related to test planning and coverage include:
  - Performing model testing as an additional activity to verify that a model fulfills its model requirements specifications (while this can be technically seen as a module/unit testing activity, it still requires close oversight by the regulated company to ensure that the model is fit for purpose for its context of use)
  - Consideration of AI-specific functionality in assurance activities such as supplier testing, unscripted testing, scripted testing, and user acceptance testing
- Consideration of additional specific test types that address AI-specific functionality and risks
- Integration of the model into the AI-enabled computerized system may depend on the product and use case
  - Note that the use of various environments such as development, formal testing, and operational environment still applies to AI-enabled systems
  - For example, a product provided by a supplier may offer the functionality to perform model testing after integration into the system, while model testing may only be feasible before integration within the MLOps architecture (see Appendix M8) in other cases
- Relevance of performance indicators that rely on data that is fit for purpose of the intended use, particularly when leveraging previous testing including supplier testing

The following subsections provide further guidance regarding key activities when testing AI: model testing, model and AI sub-system integration, and deployment and acceptance.

This guidance is intended to be used in parallel with *ISPE GAMP 5 (Second Edition)* Appendix D5, which covers testing activities performed throughout the development process regarding non-AI functionalities. See further guidance on software testing from ISO, IEEE, and ISTQB® [77, 78, 79].

Note: As stated previously, the model and AI sub-system integration may also occur before the model testing. In this case, model testing is performed within the AI-enabled computerized system in the respective formal testing environment as per guidance provided.

Since Generative AI may be prone to weaknesses such as hallucination, i.e., generating content that seems plausible at first sight but does not hold true, dedicated testing is required for Generative AI. Testing of such applications should focus on control and sufficient oversight of generated content. To this end, various concepts may be applied, with further details provided in the next subsections:

- **Evaluation methods:** Measuring the performance of LLMs by, for example, comparing answers to independently created content by SMEs
- **Defensive user interaction design:** Users should be supported in the way they conduct human oversight; this may include providing references to original sources and contextualization of such sources regarding the generated response or highlighting of key terms that are of high relevance for the process; the suitability of such designs from an end user perspective should be tested
- **Guardrails:** Regulated companies determine the effectiveness of control of Generative AI, including controls of data input and syntax as well as embedded controls regarding the factuality of the generated content in tandem with end users, as applicable

9.9.1 Model Testing

To achieve fitness for purpose of models in their context of use, a model testing approach should be planned and documented as part of the overall test strategy, specifying how the verification should be performed. While it is the regulated company's responsibility to ensure such fitness for purpose, they may choose to delegate some of the testing activities to suppliers.

The test manager is responsible for planning testing approaches and types of testing, working closely with the AI sub-system owner. Additional SMEs with domain and technical expertise provide support in assessing model testing results. Model testing is aligned with the process owner and the system owner, to ensure model testing activities are consistent with the overall test strategy of the AI-enabled computerized system and system prerequisites are satisfied.

Defining a testing approach includes:

- Choice of depth and rigor
- Types of testing activities
- Specific types of testing
- Use of data
- Acceptance criteria

Model testing is performed in a testing environment, ensuring that the use of the test data set is controlled. Model testing results need to be protected from changes and traceable with regards to models, configuration, choices of performance indicators and thresholds. The testing environment may be part of the MLOps architecture (see Appendix M8).

Information and records on testing activities should be captured. In addition, as outlined in FDA draft guidance [55], any deviations from the plan should be captured. Risks to patient safety, product quality, and data integrity should be considered as potential implications of these deviations. The rationale or further evidence should be provided to demonstrate the extent the testing results provide a meaningful basis for the determination of the fitness for purpose of the model.

9.9.1.1 Choice of Depth and Rigor

The choice of depth and rigor of model testing is based on complexity, novelty, the model risk, and residual risks (see Appendix M3). Depth relates to the size and expectations on the test data set as well as the chosen types of testing activities and specific types of testing, while rigor refers to the formality of performing model testing activities, including the assessment of model testing results.

9.9.1.2 Types of Testing Activities and Specific Types of Testing

Types of testing activities include white box and black box testing (see *ISPE GAMP 5 (Second Edition)* Section 25.6.2 [2]), with a choice of specific types of testing for AI. The choice of types of testing activities and specific types of testing depends on the nature and purpose of the model in the context of use, as well as risk.

Model testing can include the following specific types of testing:

- **Normal Case Testing:** Testing whether an exact output is provided, given the model input; the feasibility of normal case testing depends on the complexity of the model
- **Negative Testing:** Testing its behavior under unusual or extreme conditions

- **Adversarial Testing:** Testing its resilience to variations in model input
- **Repeatability Testing:** Testing the replication of data sets and model development results when repeated under identical conditions
- **Performance Testing:** Testing its adaptability to and performance on new, unseen data
- **Fairness Testing:** Testing for possible concerns regarding fairness or bias
- **Explainability Testing:** Testing its ability to provide interpretable model output
- **Reproducibility Testing:** Testing its ability to reproduce the model output when provided the same model input and configuration (e.g., through various environments); reproducibility may not be fully applicable for Generative AI (see also consistency testing below)

For Generative AI, additional specific types of testing include:

- **Consistency Testing:** Testing its ability to provide consistently correct output when provided the same input multiple times
- **Correctness Testing**: Testing its ability to create factual results

Depending on the complexity and nature of the model, structural or path testing can be performed on the structure and algorithmic functionality of the model.

When a reference model is available, regression testing may be considered to test the ability of the model to still perform as intended after changes have occurred; model output between the model and its reference are compared against thresholds.

When considering dynamic system designs, testing should also include simulating the model evolution path based on unseen data, to verify the suitability of the dynamic system design and its controls such as stop criteria for dynamic learning.

While supervised learning approaches offer a rich and established space to determine the performance of models leveraging the labels (ground truth in this case), qualitative assessment and evaluation of model results is more important when employing and leveraging unsupervised learning approaches in the absence of the ground truth. A typical case is the clustering of incident descriptions via unsupervised learning approaches, which aims at exploiting hidden structures in the data to derive a more comprehensive overview of incidents. In such situations, generated clusters and categorizations as well as the assignment of entities to these clusters should be assessed from a cross-functional perspective to determine the suitability of the model. Additionally, assessing the robustness and generalizability of the model should be considered, e.g. by assessing the behavior of the model via changes of model input (e.g., adding random noise) and the model's sensitivity in deriving model output on changed model input compared to the original model output.

9.9.1.3 Use of Data

A test data set, that is fit for purpose in the context of use of the regulated company, is required to perform model testing. When delegating testing activities to suppliers, the regulated company should consider:

- Providing data that fulfills the expectations of being fit for purpose
- Assessing the fitness for purpose of data provided by the supplier or its sub-suppliers

- A hybrid approach involving both own data and data provided by the supplier so that a fit for purpose test data set can be constructed

It is important that data originating from data splitting (see Section 9.6) and designated to the test data set remain untouched during the iterative experimentation process. In addition to data created during data splitting, additional data such as real-world data or synthetic data may be used, as applicable. The choice of using additional data depends on the intended use and the availability of data (see Appendix M8).

9.9.1.4 Model Acceptance Criteria

Model acceptance criteria originate from model requirements specifications; they are represented by a set of model performance indicators and thresholds.

The fulfillment of acceptance criteria is determined based on performance indicators derived when applying model release candidates to the test data set and, as applicable, additional data. Performance indicators also allow for a quantitative evaluation across various model release candidates and determination of the final designation among them for integration into the AI-enabled computerized system.

If the model release candidates do not meet expectations, further analysis can be performed to identify reasons for unsatisfactory performance. The assessment may include details on model testing results such as performance indicators, data quality and its potential shortcomings, data pre-processing methods, and the model design and specification. Project teams should consult further SMEs to improve the understanding of the interplay between model input, the model, and model output in the model's context of use based on this assessment. The FDA draft guidance [55] suggests various options, including:

- Downgrading model influence
- Increasing rigor of assessment or augmenting model output by using further development data
- Establishing additional controls to mitigate risks
- Changing model development approach (e.g., refining preprocessing techniques or adjusting hyperparameters)
- Rejecting the use of the model approach in the context of use, or refining the context of use

In case an iterated model release candidate is chosen, it should undergo further testing, utilizing new test data to ensure an impartial assessment. EMA states in the Reflection Paper on AI in the lifecycle of medicinal products [43] that a test set cannot be used again; rather a new and independent test data set should be employed. However, if a new test data set is not feasible, the decision to reuse data should be based on risk, particularly data leakage, and a rationale should be provided.

9.9.1.5 Release Candidate Selection

If a release candidate meets acceptance criteria, it is eligible for integration into an AI-enabled computerized system. If various model release candidates meeting acceptance criteria are selected, then a comparison and assessment of those models should be performed. Factors that contribute to identifying the most suitable model release candidates include:

- Performance of the release candidate as per model performance indicators; a balance between various performance indicators, linked to risk and the context of use, needs to be achieved
- The complexity of the model; typically, a less complex model with similar performance is favored over a more complex model

9.9.2 Model and AI Sub-System Integration

The following prerequisites need to be met to integrate an AI sub-system and its model in the AI-enabled computerized system:

- Information and records on the model release candidate's performance
- In the case of involvement of suppliers, further information on previous testing activities
- Non-AI functionality to integrate the AI sub-system into the AI system, including programming and user interfaces: specified, implemented and its fitness for purpose verified
- Established deployment procedures to instantiate the AI sub-system within the AI-enabled computerized system integrated in the formal testing environments
- Established change management procedures that allow for transition from one version of a model to the next within the AI-enabled computerized system

When integrating the AI sub-system, modularization involves isolating functionality used for development of models from productive functionality (see *ISPE GAMP 5 (Second Edition)* [2]). For instance, inference functionality to generate model output from model input may be relevant only for a static system design, while functionality used for model training and model evaluation or hyperparameter tuning should be excluded. However, such functionality becomes part of the AI sub-system when considering dynamic system design.

Planning for deployment includes the procedures for version control to track changes and enable users to revert changes.

9.9.3 Deployment and Acceptance

Provided that the AI-enabled computerized system can be deployed, a formal testing environment that aims for close mirroring of real-world use should be set up.

Once the integrated AI-enabled computerized system is deployed in the test environment, execution of initial tests can prepare further testing activities. Such tests ensure integration of the AI sub-system with the AI-enabled computerized system; for instance, smoke testing may cover the correct integration of model input by the AI sub-system within the AI-enabled computerized system, pre-processing, inference, and post-processing.

Depending on the infrastructure setup and the context of use, dedicated testbeds may be considered for testing activities. The goal is to create a promote rigor, transparency, and replicability of testing procedures, when evaluating functionality, usability, and performance – as critical input to the determination of the fitness for intended use, and as safeguard for the operation phase. See NIST AI TEVV [80].

Model testing results inform the testing strategy of the AI-enabled computerized system, with close involvement of the process owner and system owner, as well as the AI sub-system owner and test manager.

Testing activities of the AI-enabled computerized system, including User Acceptance Testing, should be performed in accordance with the plan, which may include more formal testing activities (e.g., scripted testing) or less formal testing activities (e.g., informal or exploratory testing depending on the level of risk). See Appendix M3 and *ISPE GAMP 5 (Second Edition)* [2].

Key aspects for testing activities, in support of achieving effective, fit for intended use, and compliant AI-enabled computerized systems, include:

- Performance testing, i.e., evaluating the AI-enabled system's performance under various operational conditions while ensuring that data used for testing sufficiently covers scenarios expected in operation
- Positive cases (normal case testing), i.e., situations where the AI-enabled computerized system utilizes the AI sub-system for providing model output to users or in the process
- Invalid cases (negative case testing) that reflect situations when the AI sub-system should not provide model output, if applicable
- Involvement of the AI sub-system owner in functional testing performed as acceptance testing; specific tests should be run to ensure that the AI sub-system owner can perform the oversight for ongoing model monitoring
- Involvement of SMEs and technical experts such as AI and ML engineers (which may be represented by the AI sub-system owner) to support functional testing; specific tests should be run to ensure that sufficient information is provided to perform assessments of model output, for instance in incident management scenarios
- Test procedures to derive new model versions and deployment of a new model version, including human verification of new models' performance, depending on the chosen adaptiveness as per Maturity Level; consideration of a higher infrastructure load for model engineering purposes
- Comparisons between the performance of the AI sub-system, and its model, integrated into the AI-enabled computerized system against previous model testing activities; evaluation of possible gaps in performance regarding their risk
- Ensuring that end users can apply sufficient oversight, depending on the chosen autonomy as per Maturity Level, which may involve the use of XAI methods and a test of their suitability
- Ensuring that end users provide adequate feedback, if applicable, demonstrating awareness of the implications of their actions for monitoring and further model development purposes
- Assessment of whether the AI literacy level established through education and training is sufficient to support the use case
- Verification of the traceability of data and models; this includes the relation between model input, model versions, and model output
- Resilience regarding AI-specific threats as part of performance testing; see Appendix S5
- Ability to scale effectively, to handle loads expected for use in operation without performance degradation as part of volume/load testing

In addition, testing of dynamic systems includes:

- Demonstrating control of the evolutionary path function as intended; this includes positive cases (where a new model version can be instantiated) and negative cases (where a new model version is retained, and the system is switching into static mode)
- Verification of the suitability of the dynamic system design from the end users' point of view

Testing activities should be scaled in depth and rigor based on the impact, risk, complexity and novelty, outcome of supplier assessments, the AI maturity, and the nature of the use case.

9.10 Reporting and Release

Reporting includes the preparation of a validation report as a basis for the formal determination of fitness for intended use.

Reporting should consider information and records captured throughout the AI project, based on information captured during the concept phase including requirements and as planned during planning (see Section 9.3). Elements relevant for reporting specific to the use of AI include:

- Information and records generated by risk management activities
- Information and records regarding the suitability of the IT infrastructure
- Information and records regarding the choice and preparation of data, in particular test data sets that serve model testing purposes
- Model requirements specifications
- Information on design decisions, like the determination of and rationale for the chosen model design space
- Relevant information from iterative experimentation and the selection of model release candidates
- Testing activities of the AI-enabled computerized system, including model testing and fulfillment of acceptance criteria
- Information gathered regarding supplied data, services, or AI-enabled software products

The report should also include any deviations from the plan, outstanding and corrective actions, as well as additional information gathered for the non-AI-components of the computerized system (see *ISPE GAMP 5 (Second Edition)* [2]).

If the computerized system is determined to be fit for intended use, the release completes the project.

9.10.1 Prerequisites for Release

The release of an AI-enabled computerized system involves several prerequisites to ensure the safety and effectiveness of subsequent operations. (See Appendix P3.) The following prerequisites highlight aspects that are of primary relevance for AI-enabled computerized systems, while general guidance as per *ISPE GAMP 5 (Second Edition)* [2] applies:

- Fitness for purpose of any AI sub-systems
- Qualification of infrastructure components
- Validation for the intended use of the AI-enabled computerized system
- Establishing sufficient training for end users and other stakeholders
- Preparation and setup of ongoing performance monitoring
- Fitness for purpose of functionality for management of live data
- Planning of the deployment including possible rollback plans
- Setup of continual improvement processes

9.10.2 Deployment

Deployment is performed according to the plan. Ideally, deployment is highly automated (e.g., as part of a CI/CD pipeline) and equipped with effective quality gates, as it is typically executed frequently due to the iterative nature inherent in developing and updating AI sub-systems. The exact process will differ depending on the system environment, including the use case, the AI sub-system type, the MLOps infrastructure, and the DevOps processes in place (see Appendix M8).

9.10.3 Deliverables

Key deliverables for release of the AI-enabled computerized system include the following:

- **Deployed AI-enabled computerized system:** The AI-enabled computerized system, including its AI sub-systems, is deployed in the operational environment.
- **Deployment Records:** Results of the successful execution of the deployment procedure are recorded.
- **Validation Reports:** These reports confirm that the AI-enabled computerized system in operation is fit for intended use and meets the necessary regulatory and compliance requirements. This includes determining the fitness for purpose of all AI sub-systems integrated into the AI-enabled computerized system.

10 Appendix P3 – Operation Phase

10.1 Introduction

This appendix describes processes and activities performed during the operation phase of the AI-enabled computerized system introduced in Chapter 3.

The goals of the operation phase, as stated in *ISPE GAMP 5 (Second Edition)* [2], are maintaining GxP computerized systems in a compliant state, and ensuring the systems continue to maintain patient safety, product quality, and the integrity of GxP data and records. Additionally, for AI-enabled computerized systems:

- Maintenance includes ensuring the performance and quality of models embedded in AI sub-systems
- Data in addition to GxP data needs to be considered
- There needs to be assurance of adequate human oversight, including consideration of changes in end users' behavior when acting on model output

In addition, regulated companies are expected to establish a continual improvement management system to increase quality or performance.

10.2 Overview

The AI-enabled computerized system is used in operation to support the regulated process. An overview of elements in the operation phase is shown in Figure 10.1.

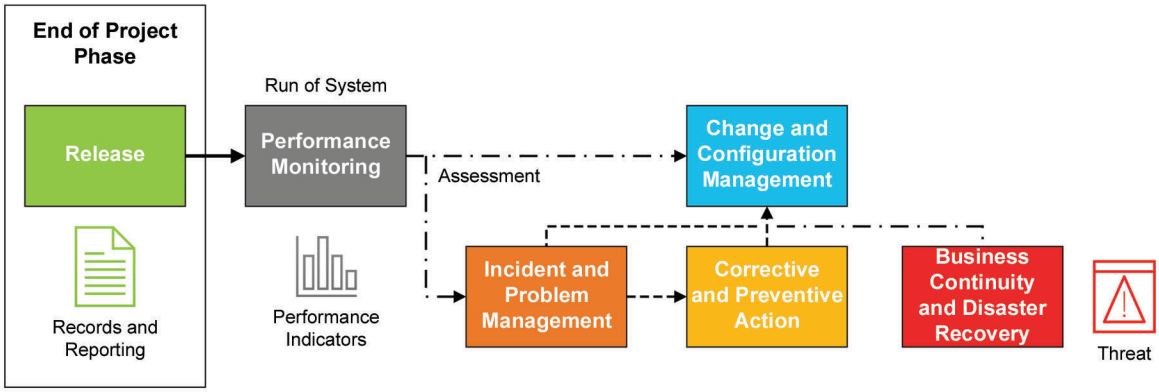
Ongoing monitoring evaluates the performance of models embedded in AI sub-systems based on defined performance indicators. User feedback is collected and assessed. It unveils opportunities for improvements, both proactive and reactive, or adapting the context of use and the scope covered. With effective knowledge management processes, changes are communicated to involved stakeholders. As changes occur and new data is collected, new knowledge is gained.

Training, fine-tuning, evaluation, and testing, as well as additional configuration or coding may be needed when processing live data. Processing live data may motivate concept phase and project phase activities to enhance the AI-enabled computerized system, such as the choice of AI maturity, training, fine-tuning, evaluation, and testing, as well as additional configuration or coding. Change and configuration management processes ensure traceability of the AI-enabled computerized system components to maintain a state of control, according to the chosen AI maturity; see Appendix M10.

These processes are especially critical when triggered by an incident or as the consequence of a CAPA, potentially requiring an update (correction) of the AI sub-system or the AI-enabled computerized system.

Finally, most activities described in this appendix for the operation and management of an AI-enabled computerized system require concise and unambiguous guidelines and instructions. Providing these (e.g., as SOPs) and ensuring adequate training helps achieve a consistently high-quality level of AI-enabled computerized systems.

Figure 10.1: Operation Phase Overview



As stated in *ISPE GAMP 5 (Second Edition)* [2]: "*Scalability should be considered [when designing operation phase processes], that is, the controls can be implemented at a level of formality and complexity appropriate to the individual organization and across a wide range of systems. Organizations may have straightforward controls for simple systems and more sophisticated tools and procedures for systems with increased impact, size, and complexity.*"

For AI-enabled computerized systems, the IT infrastructure tools and components should be appropriate to support specific requirements of AI in the context of use; refer to the concept of an MLOps architecture in Appendix M8.

10.3 Handover

"The project management approach should include a process for handover of a system from the project phase into the operational phase." [2] Relevant items include:

- Confirmation of the validated state of the AI-enabled computerized system or service
- Managed handover of open items
- Completion of data migration (as appropriate)
- Update of configuration items
- Establishing support agreements and support controls
- Knowledge transfer
- Hypercare and Monitoring of Business Readiness

Specific to AI-enabled computerized systems, communication and education involves insights from iterative experimentation and model testing results that indicate potential limitations or residual risks.

The operation team leverages monitoring processes that provide model performance compared to established benchmarks and triggers, in support of maintaining a state of control.

The chosen maturity level is clearly communicated to end users. Specifically, end users should be made aware of dynamic system designs and the potential for changing the behavior of models, if appropriate.

On a more technical level, planning and performing a handover from project change management to operational change management facilitates effective use and maintenance of the AI-enabled computerized system, following defined and documented processes. Verification of handover includes:

- Acceptance and establishment of change management and configuration management processes by the operational team, including system and process owners and technical support teams
- Review of configuration records
- Verification that backup processes and disaster recovery plans are in place and have been tested

Tools can be considered to help organize and maintain controls over software development, ensuring consistency and managing change effectively across the development and production environments; see Appendix M8.

10.4 Establishing and Managing Support Services

When establishing and managing support systems, regulated companies should:

- Identify support needs
- Assess and select providers
- Establish SLAs
- Establish support SOPs
- Monitor quality and performance

In addition to the service descriptions provided in SLAs for non-AI-enabled computerized systems, regulated companies should add AI-specific elements, including:

- Monitoring model performance compared to established benchmarks
- Data and model management activities needed to create new model versions
- Change management aspects of AI sub-systems and changes to models
- Incident response plans for issues identified in models
- Level of expertise for incident management and RCA, typically requiring data science and domain expertise

10.5 System Monitoring

System monitoring is the ongoing process of monitoring and reporting *"system failures, availability, performance, configuration baseline, and information security issues."* [2]

For AI-enabled computerized systems, system monitoring is evaluating the performance of AI sub-systems to ensure they operate effectively and efficiently over time, matching expectations measured by performance indicators against their thresholds.

Performance indicators are produced regularly or provided by real-time monitoring. Model performance and data quality should be in scope of monitored metrics and performance indicators.

If a performance indicator does not meet its requirements, modifications, retraining, or fine-tuning of models and its AI sub-system is initiated; a major redesign of the AI sub-system may be needed.

Monitoring activities, deliverables, and responsibilities should capture:

- What is monitored
- Who is responsible for the monitoring, including assurance of appropriate training
- How the monitoring is performed
- When should the monitoring occur (e.g., continuous, at each release)
- How monitoring activity results are documented, including decisions made

Dedicated strategies for monitoring should be planned for static and dynamic system designs:

- For static system designs – monitoring model performance indicators and data characteristics of the series of deployed model versions per change and configuration management
- For dynamic system designs – monitoring adapting models in response to evolving conditions and data distributions

10.5.1 Monitoring Static AI Sub-Systems

Established procedures cover monitoring and notifications for static AI sub-systems to detect performance issues and deviations:

- **Data quality checks:** Checks for data quality cover missing values, outliers, edge cases, or changes in data distributions, as they may affect an AI-enabled computerized system's reliability. Live data is compared to data used for model development and testing to demonstrate representativeness.
- **Anomaly detection:** Processes for anomaly detection aim for detection of unusual or unexpected patterns in data.
- **Comparison to a ground truth or reference standard**: Deriving the ground truth, or reference standard, can be intrinsic to the process or established through methods such as random (manual) verification. Effort in collecting data representing the ground truth or reference standards should be commensurate with the level of risk. Assumptions built in AI sub-systems are evaluated and verified that their models stay fit for purpose in the context of use.
- **Bias monitoring:** Bias monitoring assesses model output for biases across different subgroups of model input. Relevant subgroups of data are identified that are informative to address sub areas within the context of use and potential challenges for fairness.
- **Analyzing performance indicators:** Statistical methods and visualization techniques help analyze the performance of models and detect potential issues.
- **End-user feedback:** Monitoring and assessment of end-user feedback covers interpretability of models and effectiveness of XAI methods, as well as meaningful interaction with the system.

Thresholds on performance indicators trigger alerts when expectations are not met, enabling timely responses to potential issues.

10.5.2 Monitoring Dynamic Systems

The monitoring process for dynamic systems extends that of static systems, considering aspects rooted in the higher adaptiveness of the system; see Appendix M10:

- **Monitoring evolving models:** Comparing performance indicators of more recent and past model versions detects unexpected changes and shifts in performance and model behavior.
- **End-user feedback to adapting models:** End-user feedback regarding suitability of evolving model versions provides insights into the acceptance of model changes.
- **Switches to static mode:** When limits and boundaries of the dynamic system are reached, the model switches to static mode to initiate further assessment.

10.6 Incident Management and Problem Management

> *"An incident relates to the effect of an unplanned interruption to a service, or reduction in service quality, typically linked to a breach of the SLA, user observation, or feedback from automated monitoring tools. A problem relates to the root cause of one or more incidents. Problems can be raised in response to a single significant incident or multiple related incidents. Problems are the cause and incidents are the effect."* [2]

Incident and problem management processes need to allow for the potential for highly complex AI-enabled computerized systems, which can result in:

- Bias in data
- Model drift
- Unexpected behavior
- Ineffective human-AI interaction
- Security breaches

Incidents and root causes need to be identified, evaluated, and resolved to maintain the effective and safe use of the AI-enabled computerized system. CAPAs can be a consequence of such activities, as discussed in Section 10.7.

AI Incidents versus AI Hazards

ISPE GAMP 5 (Second Edition) [2] and many other publications use the term "incident." However, some publications distinguish between AI *incidents* and AI *hazards.* The Organisation for Economic Co-operation and Development (OECD) [81] defines an **AI incident** as *"an event, circumstance or series of events where the development, use or malfunction of one or more AI systems directly or indirectly leads to any of the following harms…"* and an **AI hazard** as *"an event, circumstance or series of events where the development, use or malfunction of one or more AI systems could plausibly lead to an AI incident, i.e., any of the following harms…"* The harms listed by the OECD paper include injury or harm to persons or groups of people, disruption of operations, violations of human rights, or harm to property, communities, or the environment.

In this ISPE AI Guide, "incident" is used in the operation phase, while "hazard" is addressed in risk management Appendix M3.

10.6.1 Objectives and Prerequisites

Incident and problem management processes help to maintain the AI-enabled system's fitness for its intended use and ensure it is in a state of control. Effective incident and problem management relies on:

- **System monitoring:** System monitoring may trigger incident and problem management processes if expectations on data or model performance are not met (Section 10.5).
- **Periodic review:** Periodic reviews may unveil inconsistencies or unexpected shifts, thus leading to incident and problem management activities (Section 10.9).
- **CAPAs:** Problems and incidents may result in CAPAs (Section 10.7).
- **Change management:** Processes for change management allow for configuration updates to resolve issues and problems (Section 10.8).

Operation teams performing incident and problem management activities should possess sufficient AI literacy and an understanding of the AI-enabled computerized system, its AI sub-systems, and relevant regulations. The information and records that can support operation teams include:

- **AI-enabled computerized system architecture overview:** The overview of the system including its components and AI sub-systems.
- **AI sub-system architecture:** Structural elements including data, model types and hyperparameters.
- **AI sub-system maturity:** Information on the chosen adaptiveness and autonomy.
- **Model parameter:** Parameters configured or derived from training; whether they are informative or not depends on the complexity of the model.
- **XAI methods:** XAI methods provide insights into features and drivers of the model output to understand possible shortcomings in the model; see Appendix S4.
- **Dependencies:** The dependencies needed to run the AI sub-system (e.g., packages or libraries).
- **Data traceability:** The data sets used to train (if applicable), evaluate, and test the model. When using pre-trained models, record the location from which the model was acquired.
- **Model generation code:** The code that specifies how the model was created, including techniques for data splitting, data engineering, and model engineering.
- **Performance indicators:** Assessment of model performance indicators, and their changes over time.
- **Knowledge database:** Knowledge gained throughout the system's life cycle, including articles, best practices, insights, and effective solutions for the AI sub-system should be collected, managed, and made available [82].
- **Known Error Database (KEDB) records:** Knowledge article that lists the known errors or issues within the AI sub-system. The KEDB should include a concise description of the error, its root cause, and a step-by-step resolution plan.
- **Robust IT infrastructure and integration capabilities:** Capabilities to support the computational requirements of ML algorithms, IT systems, software applications, and IT infrastructure components (servers, databases, networks, cloud services, etc.).

- **Service Level Agreements (SLAs) and Operational Level Agreements (OLAs)** defined for suppliers.

10.6.2 Implementing Incident Management Processes

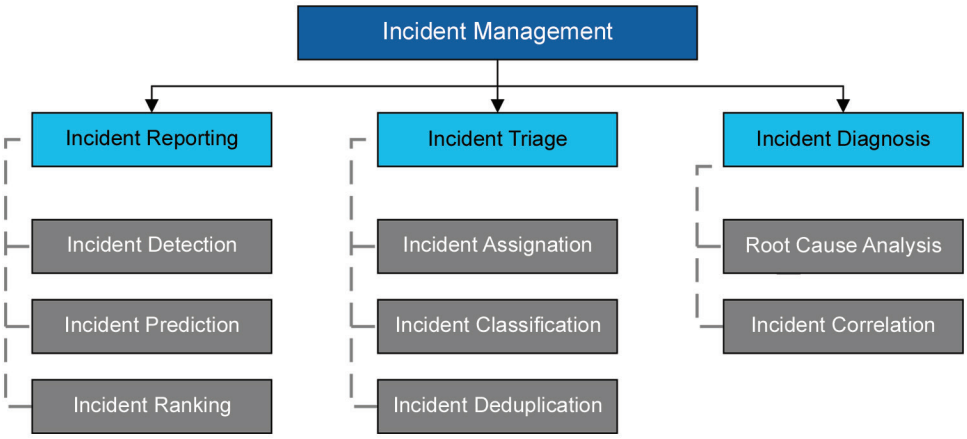
Incident detection is the first activity in the incident and problem management process (see Figure 10.2). Monitoring systems support oversight and performance control using metrics established during the project phase.

Reporting incidents captures information on the AI sub-system's state and the model input and model output linked to the incident. Incident reporting channels, such as incident ticketing systems, logging frameworks, and automated incident reporting systems can help facilitate effective communication and collaboration across operation teams and other stakeholders.

Once detected and reported, the incident is categorized using an incident triage process. This process classifies and prioritizes incidents by severity, impact, and urgency using criteria typically based on the impact to the system's functionality and business criticality.

If an incident is recurrent, an incident diagnosis helps identify underlying factors that lead to issues, errors, or anomalies. ML may be employed to identify patterns and correlations that lead to incidents.

Figure 10.2: Incident Management



10.6.3 Results and Outcomes

The incident and problem management process typically results in:

- Incident logs and records
- Reporting results of RCA
- (Updated) Risk assessments
- Incident metrics
- Updated KEDB records
- Training materials and knowledge articles
- Change control activities and records

10.6.4 Typical Incident Scenarios

Typical incident scenarios, shown in Figure 10.3, include:

1. **Deployment of the Wrong Model Version:** A redeployment with the correct version of the model is initiated, following standard deployment procedures, to correct this incident. In case this is not feasible in the required time, the correction may be switching to a formerly stable and robust version or switching to processing without the support of the affected AI sub-system provided organizational requirements, for example, adequate training, are met.

   Furthermore, systematic biases in the data pool need to be evaluated for their impact on further life cycle activities, such as monitoring, retraining, or fine-tuning.

2. **Integration Error of Model Output:** In case the output of the AI sub-system (e.g., model results) or further functionality (e.g., deriving explanations on the model behavior) is incorrectly integrated with the AI-enabled computerized system, a correction is initiated to deploy a new version of the system according to standard procedures.

   Potential downstream effects of that error need to be assessed (e.g., bias in monitoring activities) to determine the appropriate handling of that data.

3. **Integration Error of Model Input:** Similar to integration errors of model output, the model input errors are addressed via redeployment of correct integration code as part of the AI-enabled computerized system, following standard procedures. Implications on the case data set should be evaluated to avoid bias and inconsistencies for future development purposes.

   Data may be corrected or excluded from future development purposes based on the results of the evaluation. While the recalculation of model output typically requires more effort, there is less risk of a lack of representativeness of the collected data following this approach; therefore, a risk-based decision should be taken on the selected approach.

4. **Insufficient Performance:** If the model is not able to maintain the quality expectations during operation, various measures may be considered:

   - Limiting the applicable area within the AI sub-system's context of use or increasing the level of human oversight and control in case the performance expectations are not met in a clearly defined area of input data.
   - Retraining or fine-tuning on a larger data set including more recent observations.
   - Performing a full AI sub-system revision using insights from KPIs that indicate insufficient performance to guide the revision activities to target weaknesses.

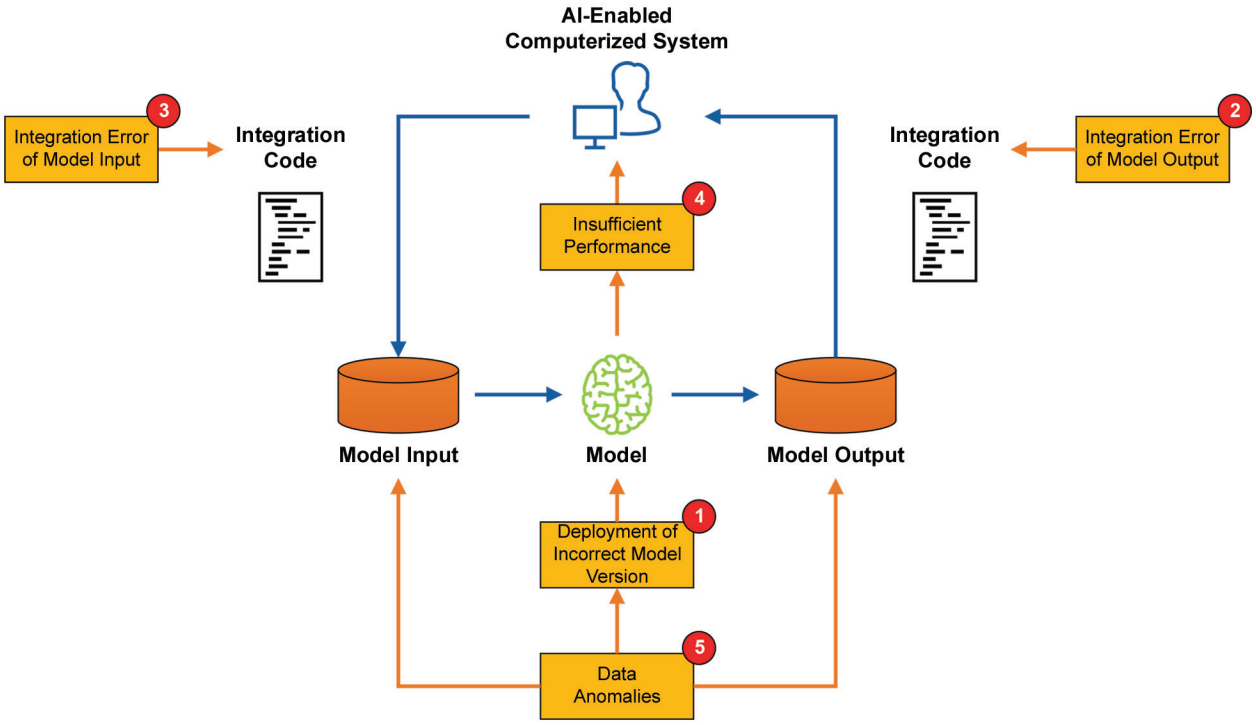
5. **Data Anomalies:** In the event of data anomalies, such as outliers or unexpected statistical distributions in the input data, implications for engineering, evaluation, and testing the model during the project phase may arise. This could raise concerns about the overall quality of the model and its output.

   Accordingly, an evaluation of the impact of the anomalies should be conducted. Consider rerunning the model engineering that led to the chosen model using cleaned or selected data to determine if adjustments to the deployed model version are necessary. In rare instances, a regulated company may determine that the selected model design is inadequate, possibly necessitating complete redevelopment of the AI sub-system.[14]

[14] An option for handling of data anomalies is "machine unlearning." This involves a specific model revision where the influence of a particular data set (the "forget set") is removed from the model. Implementations of this approach can range from dedicated retraining on the restricted data set to model updates. To safely apply such a technique, the method of unlearning needs to be understood, linked to risk assessment, and its outcome verified using suitable testing strategies, see Appendix P2. Machine unlearning activities should be performed via change control and capturing relevant information.

This incident type showcases the relevance and importance of careful data management and building data understanding throughout the model development process in the project phase.

Figure 10.3: Typical Incident Scenarios



All incidents may impact future monitoring activities of the AI sub-system; for instance, model outputs may be generated with the corrected model version in case 1 (deployment of wrong model version) to establish an adequate baseline performance to allow for identification of performance trends.

Communication of incidents should be carefully planned, involving process or senior management as applicable, so that appropriate onward action is taken (e.g., regulatory reporting, notification of other teams and stakeholders).

10.7 CAPA

"CAPA is a process for investigating, understanding, and correcting deviations and nonconformities to address the immediate impact of the issue and to minimize the risk of recurrence." [2]

A robust and effective CAPA process helps to identify root causes, evaluate problems, and initiate timely corrections, thereby ensuring the health and safety of the AI-enabled computerized system. Furthermore, by preventing the recurrence of issues, the CAPA process facilitates continual improvement.

10.7.1 Objectives and Prerequisites

As a complement to incident and problem management, a combination of general and AI-specific CAPA processes help to keep the AI-enabled computerized system fit for its intended use and maintained in a controlled and validated state throughout its life cycle.

Change management, ongoing monitoring, and periodic review activities are necessary to planning and executing CAPAs.

10.7.2 Handling CAPA

When the initial investigation is completed and the impact of an incident is known with sufficient confidence, a decision is taken if and how to address the incident and its root cause. During this step, CAPA management activities should be considered, and the creation of a CAPA plan should be evaluated. Both the activities and the plan help to apply corrective actions in a timely manner, address root causes, and prevent recurrence.

Key elements of CAPA management include:

- **Corrective action** is the reactive process of CAPA management aiming to address incidents as they occur. CAs follow the investigation of the problem to determine the impact of the AI sub-system on patient safety, product quality, and data integrity, defining the actions needed to promptly address and correct the issue (e.g., model retraining, disaster recovery, reverting to previous model version – following a risk-based approach).

- **RCA** is used to determine the root cause of the issue; it is essential to avoid only removing the symptoms. Common causes of incidents include model performance degrading over time, changing features in data that may impact model accuracy, and model bias.

- **Preventive action** is the proactive process of CAPA management that addresses problems or issues with the AI sub-system that may occur in the future. PAs use trends and recurring issues to identify factors that could lead to problems in the AI sub-system and apply measures to prevent the problems. Historical data from related AI sub-systems may be informative.

- **Analysis** of incidents and CAPA helps identify trends, recurring issues, and areas of improvement. Lessons learned from incidents and CAPA activities feed future decision-making and planning.

10.7.3 CAPA Deliverables

The CAPA process typically leads to the following information and records:

- Reporting results of RCA
- (Updated) Risk assessments
- CAPA plans and records
- Updated KEDB records
- Training materials and knowledge articles
- Change control activities and records

10.8 Operational Change and Configuration Management

Per *ISPE GAMP 5 (Second Edition)* [2], *"change management is the process by which changes to configuration items…are managed from their inception to completion."*

Change management processes cover software, configuration, data, and model changes to maintain the proper functioning and state of control of the AI-enabled computerized system.

Various drivers for changes exist. Product or quality management is looking for improvements, such as raising the efficiency of the process, elevating the system's autonomy, and extending the scope of applicability within the context of use. Other changes may be based on the results of incident and problem management activities, organizational or psychological aspects, or changing environment conditions.

"Configuration management comprises the activities necessary to define a computerized system…at any point during its life cycle." [2]

10.8.1 Goals

The design of operational change and configuration management processes depends on ongoing monitoring, risk management, cybersecurity management, Quality oversight, and AI maturity.

A change control plan needs to be established to provide guidance and promote transparency in the anticipated activities and deliverables within the operation phase, while allowing for planned changes and upgrades to AI-enabled computerized systems and their AI sub-systems.

Change control aims to maintain system reliability, performance, and accuracy and supports the following goals (see Figure 10.4):

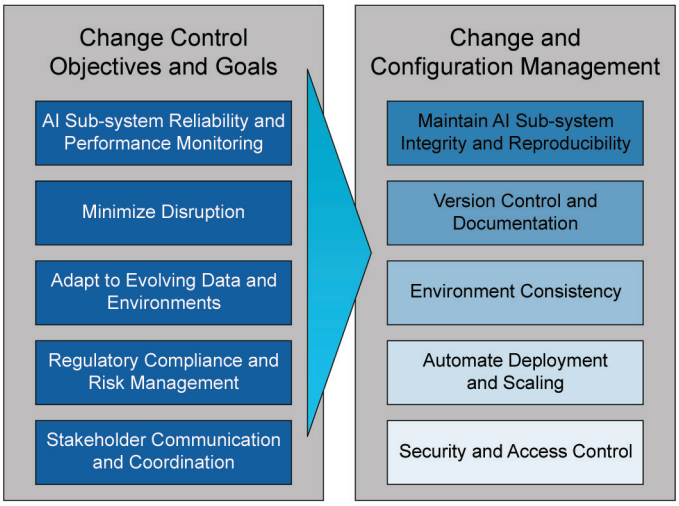
- **Ensure AI sub-system reliability and performance:** Manage the implementation of updates or changes to AI sub-systems to maintain or improve their reliability and performance without adversely impacting patient safety, product quality, and data integrity.
- **Minimize disruption:** Implement changes in a way that minimizes disruption to ongoing operations, ensuring that AI sub-systems exhibit minimal downtime.
- **Adapt to evolving data and environments:** Reflect changes in underlying data patterns, operational environments, or business requirements in AI sub-systems, ensuring that their AI-enabled computerized system remain safe and effective in supporting the business process.
- **Regulatory compliance and risk management:** Ensure that changes adhere to regulatory requirements and ethical considerations, and manage risks associated with changes to the AI sub-system.
- **Stakeholder communication and coordination:** Facilitate effective communication and coordination among stakeholders during the change process to ensure alignment and understanding.

Change management and configuration management are closely related. When the need for a change occurs, both activities need to be considered in parallel. [2]

The goals of configuration management are (see Figure 10.4):

- **Maintain AI sub-systems' integrity and reproducibility:** Maintain accurate records of AI sub-system configurations, including data sources, algorithms, parameters, and versioning, to ensure AI sub-systems can be reproduced or rolled back if necessary.
- **Version control and documentation:** Manage versions of AI sub-systems and their components to track changes over time, facilitating troubleshooting, audits, and knowledge sharing; a model history file may be established to allow for traceability of model versions.
- **Environment consistency:** Ensure consistency across development, testing, and operational environments to prevent discrepancies in evidence generated for decision-making to ensure that AI sub-systems perform as expected in operation.
- **Automate deployment and scaling:** Integrate configuration with automation of deployment to support effective scaling and updating of AI sub-systems.
- **Security and access control:** Manage access to AI sub-system configurations and deployment environments to protect sensitive data and IP, ensuring that only authorized personnel can make changes and reducing cybersecurity risk exposure.

Figure 10.4: Goals of Change and Configuration Management



10.8.2 Activities

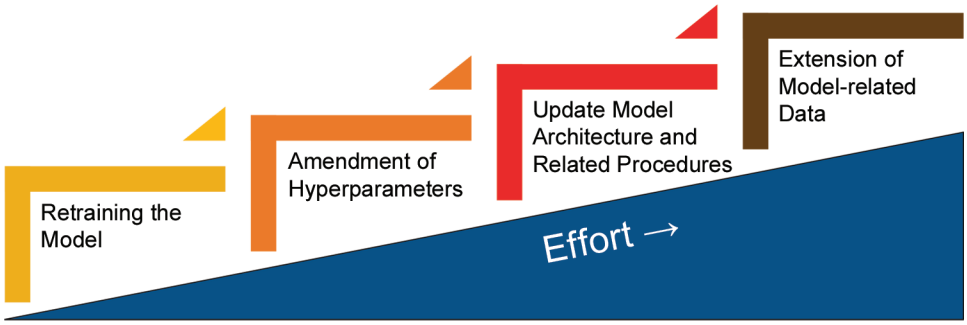
Regulated companies should ensure that changes comply with the organization's change management process. User stories and system controls form a solid base to start the change process. Drafting visuals of the process with possible scenarios and mitigations helps identify weak points [2].

The need to change to the software (or system or algorithm) can arise through multiple channels, such as:

- Customer complaints
- User complaints or findings
- Modification to computerized system use and performance (e.g., for use within a specific subpopulation)
- Feature updates
- Supplier activities
- Patch deployment
- Changes in integrated computerized systems
- CAPA findings
- New data
- Drift in performance found during KPI monitoring
- Potential improvements to performance

The change request or change proposal is the initial definition of the change outlining the business or technical reason for the change and its general scope. The typical effort and impact of changes to the AI-enabled computerized system and their AI sub-systems differ by change type; see Figure 10.5.

Figure 10.5: Typical Effort and Impact of Different Changes



Changes in AI maturity levels may be considered (Appendix M10):

- **Change in Autonomy Level:** Increase Autonomy Level (if more confident and less granular human oversight is required) or reduce (if expectations do not materialize and request tighter control)
- **Change in Adaptiveness Level:** Increase Adaptiveness Level (if need is present, i.e., regular, well-controlled AI sub-system updates are possible) or decrease (if tighter control is needed in the learning behavior of the AI sub-system).

Updates to static models should follow standard change management procedures [2]:

- **Change Request and Change Proposal and Evaluation:** *"The change request or change proposal is the initial definition of the change outlining the business and/or technical reason for the change and the general scope of the change."*
- **Change Specification** defines the change in terms of impacted configuration items (e.g., data input, functionality, configuration). It may be defined in the change record or supported by system life cycle records.
- **Impact and Risk Assessment:** *"Risk assessments should determine whether the change impacts patient safety, product quality, or data integrity. Existing risk assessment records should be updated to reflect the change. The output of the risk assessment influences the change plan and the change verification approach."*
- **Change Plan**[15] should define actions and tasks required to implement the change. It should address the change specification and the output from the impact and risk assessment.
- **Back-out Plan:** Should be considered for complex changes or those where major implications for patient safety, product quality, or data integrity are expected based on the risk assessment. It defines how the AI sub-system is reverted to a known state if the change has an unexpected impact.
- **Authorize Implementation:** *"The change should be authorized prior to implementation to confirm the adequacy of change specifications, impact and risk assessments, change plans, and backout plans as appropriate. Once authorized, the change plan can be implemented."*
- **Execute Change Plan:** Actions and tasks defined in the change plan are executed.

[15] Local guidelines and rules for specific requirements are relevant when performing change and configuration management. For example, the change management process can be part of a Predetermined Change Control Plan (PCCP) included in the device marketing submission to the FDA. [83] This PCCP allows for intended modifications to the AI sub-system to be implemented automatically (i.e., a dynamic system) or manually (i.e., requiring human input, action, review, and decision) without revising the FDA filing.

- **Release to Operations:** Once the change has been verified and change records reviewed to confirm all actions in the change plan have been fully executed in accordance with the plan, the change is transitioned back to operation (i.e., operational environment). Knowledge transfer to the support organization should be completed before the transition to operation.

Considering dynamic systems, the system is allowed to perform updates within a certain range, guarded by controls (see Appendix P2). As models in dynamic systems evolve, adequacy of the level of detail captured in specification and risk assessment should be maintained. There should be a balance between a general description of an applicable model space (in which the model may evolve) and the level of detail on specifics of the model (which may soon be outdated along the evolutionary path).

In cases where the limitations of the AI sub-system were hit and a switch to a static mode occurred, the AI sub-system and the model should be investigated timely (commensurate with the risks involved). Re-initiation of dynamic learning is seen as a change as well; therefore, records on the information gained during the investigation and the rationale for re-establishing dynamic learning should be captured.

Substantial changes, for example, increases in the maturity level, or major extensions, may require concept phase or project phase activities, typically conducted in parallel to operating the system, necessitating integration of development and operational activities.

Typical configuration activities, according to *ISPE GAMP 5 (Second Edition)* [2], include:

- **Configuration Identification:** What to keep under control

  The system should be broken down into configuration items. These include business processes and steps, software products, components, modules, infrastructure, models, and data.

  "The list of configuration items and their status (e.g., version) is called the configuration baseline and serves as reference for the validated state of a system."

- **Configuration Control:** How to perform the control

  "Changes to business processes, software, hardware, data, infrastructure components should be managed in accordance with risk and should only be made by authorized personnel" following an approved change control plan. "Configuration records should reflect the implemented changes."

- **Configuration Status Accountability:** How to document the control

  "Configuration records should be traceable to the relevant configuration item and version. As a minimum, configuration records should be traceable to the overall system release."

- **Configuration Evaluation:** How to verify that control

  "Documentation and records defining the configuration baseline should be subject to appropriate document management and/or records management controls. Configuration records should be maintained in an as built state. Configuration records are subject to records retention as defined by company records retention policies."

Note: Ensuring traceability of the data used for model development, model testing, and testing the AI-enabled computerized system (see Appendix P2) may serve a similar purpose as configuration management described above. The scope of establishing such traceability varies across systems and the level of external suppliers' involvement.

"Periodic review of operational systems should ensure that the configuration status of defined configuration items is maintained in line with changes." [2]

The records and information that provide evidence on the effectiveness of change control allowing for traceability and support in achieving compliance include:

- Retraining and reverification records
- Updated associated documentation (e.g., specification, requirements, risk assessments)
- Configuration records (e.g., specifications, data sets, tools (configuration management tools, discovery tools))

10.9 Periodic Review

> *"Periodic reviews are conducted throughout the operational life of a computerized system to verify that it remains in a validated state, complies with current regulatory requirements, is fit for intended use, and satisfies company policies and procedures."* [2]

Specific to AI-enabled computerized systems, periodic review activities allow for assessing the larger picture of incremental changes in the system or data and performance drifts, while ensuring overall consistency of information and records.

Periodic review activities, deliverables, and responsibilities should be defined:

- What is reviewed
- Who is performing the periodic review
- How the periodic review is performed
- When should the periodic review occur (e.g., after x number of changes to the sub-system)
- How the periodic review's results are captured, reviewed, and assessed, and required mitigation decisions based on the review are taken

The following factors influence the development and design of periodic review processes:

- Risk
- Duration of operation of the system
- Experience of the organization in employing models in the context of use
- Human capabilities (e.g., education, training, change in skill base)
- Data, including iterations of data splits
- Performance indicators and their changes over time
- Combined effect of various, incremental changes of models
- Choices on the maturity level of AI sub-systems
- Insights from change, incident, and problem management
- Supplier (e.g., change in supplier, patch assessments)

- Integration into the wider process and IT landscape
- Environment, social, and governance aspects, including changes in statues and regulations

10.10 Business Continuity Management

Business continuity management and disaster recovery are proactive strategies that mitigate or reduce the impact a catastrophic event might have on an organization's ability to reliably deliver its products and services.

- **Business continuity** outlines how a business will proceed during and following a failure or disruption by providing contingency plans. Planning should also address smaller interruptions or minor disasters, such as extended power outages.
- **Disaster recovery** refers to planning the response to a catastrophic event, such as a natural disaster, fire, or cybercrime. It involves the measures a business takes to respond to an event and return to safe, normal operation as quickly as possible.

Business continuity focuses on keeping the operation running, while disaster recovery focuses on restoring data access, IT infrastructure, and systems after a catastrophic event has occurred. They also have different goals, as business continuity plans limit operational downtime, whereas effective disaster recovery plans limit abnormal or inefficient system functions. Comprehensive preparation for disastrous events requires the combination of both disciplines.

Business continuity and disaster recovery for AI-enabled computerized systems generally adhere to the same practices used for non-AI enabled computerized systems with the addition of:

- Ensuring the quality of data
- Ensuring the integrity of AI sub-systems and models
- Determining the impact on model performance

Careful planning and preparation in previous life cycle phases are required for successful business continuity management and disaster recovery planning. Both require diligent data, model, and system management, including thorough traceability and version control of all relevant artifacts and deliverables; see Appendix M7.

While basing decisions and actions on preplanned activities and procedures, each situation requires careful assessment regarding the suitability of plans and assumptions taken during planning, to adequately address the actual incident that occurred.

10.11 Security Management

Per *ISPE GAMP 5 (Second Edition)* [2], *"security management is the organizational, process, and technical considerations that ensure the confidentiality, integrity, and availability of an organization's regulated computerized systems, data, and records."*

An overview of key security measures and references to standards are described in *ISPE GAMP 5 (Second Edition)* Appendix O11 [2]. In this ISPE AI Guide, Appendix S5 discusses measures specific to AI-enabled computerized systems.

11 Appendix P4 – Retirement Phase

11.1 Introduction

Regulated companies may choose to retire an AI-enabled computerized system or one or more of its AI sub-systems for various reasons, including poor user experience, changes in regulatory requirements, associated costs, availability of better solutions, inadequate performance, or end of support from the supplier.

This appendix assumes that the regulated company has decided to retire an AI-enabled computerized system or sub-system and has identified the data to be archived, with possible migration to other system(s), as applicable.

Once the organization understands that the system or the AI sub-system needs to be removed or replaced, the following options can be considered:

- **Retirement:** The AI-enabled computerized system or any of its AI sub-systems is removed from active operations, i.e., normal operational users are deactivated and interfaces are disabled. Models in the scope of retired (sub-)systems are not used further. Restricted access is maintained for any regulatory needs, data reporting, results analysis, and support.
- **Decommissioning:** The controlled shutdown of a retired system. An AI-enabled computerized system may be stored if required to be reactivated later, e.g., for retrieval of regulatory data or results.
- **Disposal:** Data, information and records, software, or hardware are permanently destroyed. Each may reach this stage at a different time. Data, information, and records may not be disposed of until they have reached the end of the record retention period as specified in the record retention policy. [2] The same applies to models, in the case of AI-enabled computerized systems, and associated information and records such as performance indicators.

This appendix is intended to be used in parallel with *ISPE GAMP 5 (Second Edition)* Appendix M10 [2]. All aspects of system retirement are covered, focusing on aspects specific to the use of AI:

- Retirement planning
- Retirement execution
- Retirement reporting
- Lessons learned and improvements

Considerations on data and models, as well as their relations, are of high relevance for AI-enabled systems, as elaborated in the following subsections.

See *ISPE GAMP 5 (Second Edition)* Appendix M10 for general considerations for the retirement of systems. See *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* Section 3.6 [1] for details related to data.

11.2 Retirement Planning

The extent and rigor of planning should be based on the AI-enabled computerized system's impact and risks associated with ensuring data integrity and privacy.

The retirement process should be planned, with input typically from cross-functional members and approval by the process owner, system owner, Quality Unit, and others as required (such as legal representative, clients/operations representative, etc.).

AI-specific aspects that are relevant for retirement planning of AI-enabled computerized systems or AI sub-systems are highlighted as follows:

- The linkage between models and their configurations and hyperparameters, model input, model output, and performance indicators should be considered throughout the AI sub-system life cycle, including model development, model testing, testing of the AI-enabled computerized system, as well as operational use, as applicable. Similarly, human input such as verification of correction or feedback should be captured, depending on the maturity level.

- Data and models may be used in more than one computerized system; therefore, an overview of dependencies to other systems should be established or reviewed for retirement planning purposes.

- Further and secondary use of data and models should be planned for, pertaining to restrictions such as privacy and data protection; this is linked to planning migration, archival, or destruction activities.

- Specific roles in the context of AI-enabled systems, such as data scientists and ML and AI engineers, provide support in planning in relation to dependencies of the AI sub-system within the AI-enabled computerized system.

Planning and contractual agreements for continued support from suppliers during retention periods is of high importance, including the above considerations.

11.3 Retirement Execution

General provisions for executing retirement activities apply, with consideration of additional planning aspects outlined in Appendix P4. The main aspects for retirement execution include:

- Availability of all resources before the retirement plan is executed

- Careful planning of the timing, including transition and migration to a replacement system, as applicable

- Migration or disposal of sensitive data in a secure manner, while considering further or secondary use as planned

- Business continuity plans and restoration activities planned for cases where problems arise during the retirement or migration process, including the choice of model versions, specifically for dynamic systems

- Information and records on all planned and any unplanned activities occurring during system retirement

11.4 Retirement Reporting

After the retirement plan is executed, a summary should be created to describe the execution and results, as well as the testing and verification activities including any deviations, as applicable. Additional considerations may apply in the context of AI-enabled computerized systems, including:

- Assurance of traceability of model input, models and model output, and adequacy for *ex post* assessment purposes, including the possibility to assess model and data drift
- Successful transition of the AI sub-system to the replacement system
- Successful migration of data for further or secondary use, in line with privacy and data protection expectations

It should be ensured that retirement activities were performed in a safe, traceable, and comprehensive manner as planned and meeting acceptance criteria. Upon completion, relevant stakeholders should be informed about the successful retirement activity.

11.5 Lessons Learned and Improvements

Regulated companies are encouraged to facilitate opportunities for reflection on the process and the life cycle approach, sharing experiences, and striving for continual improvements.

Consult with all relevant stakeholders to draw conclusions on:

- Positive outcomes and observations
- Negative outcomes or deviations from planned activities
- Aspects that need to be improved

These activities can foster the building of knowledge within the organization concerned with AI-enabled computerized systems; see Appendix M5. These insights can help inform change management processes concerned with life cycle management procedures and policies. Refer to CMMI [10] which provides an approach based on a framework for assessing and improving organizational capability and maturity.

12 Appendix M1 – Quality by Design (QbD)

12.1 Introduction

All computerized systems, including those using AI and their components, should safeguard patient safety, product quality, and data integrity. Additionally, appropriately managing QbD can enable the regulatory compliance of AI-enabled systems integrated into established QMS activities, per ICH Q8 [84] and expanded upon in *ISPE GAMP 5 (Second Edition)* [2].

This appendix outlines key considerations on QbD principles in the context of AI-enabled computerized systems. It focuses on AI-specific aspects of QbD, while further considerations per *ISPE GAMP 5 (Second Edition)* apply to all computerized systems.

While general aspects of the quality system are described in this appendix, further information is in Appendix S6 for medical devices.

12.2 General Considerations

For AI-enabled systems, these QbD aspects have elevated significance:

- Systematic evaluation and understanding of the process and key goals in the context of use
  - Complexity of model input
  - Model architecture and model complexity
  - Adaptiveness and autonomy design based on product and process understanding
  - Human-AI interaction considering the capabilities and limitations both of humans and AI
- Establishing a scientific based, forward-looking QRM approach
  - Interplay of data, models, and users in the context of use
  - Data quality and implications of data quality shortcomings
  - Integration of ongoing monitoring and QRM
- Proposal for design spaces that fit the process and its goals, including the definition of key performance and quality indicators
  - Range of model development choices per model design space
  - Choice of performance indicators
  - Choice of thresholds
- Use of experimentation to identify the relationship between inputs and process data on KPIs

- Compromises between competing objectives
- Comprehensive reasoning for model selection, considering the context of use

- Thorough testing to provide evidence of fitness for purpose prior to operation
  - Coverage of representative scenarios for model testing and testing of the AI-enabled computerized system
  - Assurance of effectiveness of human oversight and control
  - Rigor of records and information captured during testing
  - Adequacy of change management aspects, such as processes to signal, initiate updates, and deploy new model versions

12.2.1 Model Quality Dimensions

Due to the statistical nature of models, teams utilize quantitative measures to:

- Develop product and process understanding and data understanding
- Apply science-based risk management
- Define quality attributes
- Guide iterations during experimentation
- Determine the overall assessment of their fitness for purpose
- Perform ongoing monitoring during operations

To support QbD principles, five model quality dimensions (see Figure 12.1)[16] serve as an orientation:

- **Fit for purpose data:** A model is only as good as the data used to develop it. While the data's fitness for purpose is determined by means of relevance, representativeness, abundance, and reliability, more detailed data quality dimensions are considered good practice. Details on the concept of data's fitness for purpose are provided in Appendix M6, expanding on the significance of data and providing guidance regarding assurance of data's fitness for purpose in AI-enabled computerized systems.

- **Stability and robustness:** Generally, stability refers to the property that minor changes to an input result in minor changes to model output, which are easier to control and use in regulated processes regarding interpretability and trust, as well as downstream implications for other processes and systems. This may also support robustness against potential inaccuracies in data provision. Requirements for stability and robustness should be balanced against potential limitations in precision in case discontinuities are present in the real-world relationships between model input and model output. Another aspect of robustness is the system design resistance to cybersecurity threats.

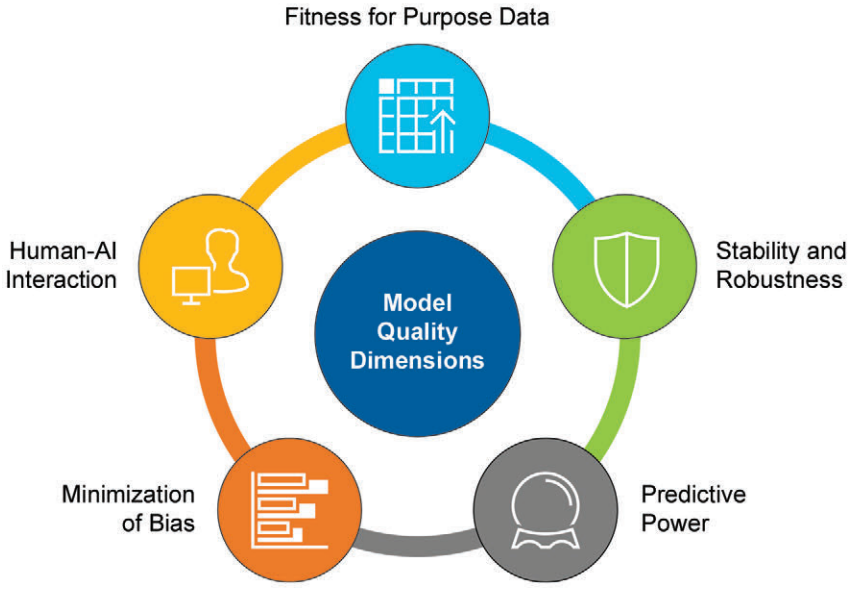
- **Predictive power:** The output of a model can be interpreted as a statistical prediction given the model input provided. Predictive power describes how far the model can reliably predict or determine the real-world outcome or value, measured by KPIs.

[16] Also, AI Governance and QA Framework: AI Governance Process Design [34], where the quality dimensions were originally established.

- **Minimization of bias:** Minimization of bias aims for a close match between model output and ground truth, concerned with a.) the entire scope within the context of use, and b.) considering subgroups and subpopulations within this scope. Minimization of bias on the overall scope aims to facilitate effective AI-enabled computerized systems. In addition, minimization of bias in dedicated subgroups can promote fair and responsible use of AI. Bias is often measured by performance indicators applied to data representing subgroups, which allows for comparison across groups.

- **Human-AI interaction:** Models are embedded in an AI-enabled system with multiple human touchpoints, ranging from active verification or interference by exception to RCA, monitoring, and periodic review activities. Therefore, the performance of the human-AI team should be assessed, including the effectiveness of XAI methods and evaluation of human user feedback; see Appendix S4.

Figure 12.1: Model Quality Dimensions



12.2.2 Integration Into the QMS

Four pillars of the QMS serve as guidance to apply model quality dimensions in an AI use case (see EMA's ICH guideline Q10 on pharmaceutical quality system [82]): process performance monitoring, problem and incident management, change management, and management review of process performance (shown in Figure 12.2):

- **Process performance monitoring:** Monitoring performance is needed to maintain a state of control. Monitoring allows for the identification of improvements or deviations to initiate change management processes and problem and incident management processes; see Appendix P3. Collection of model input, model output, and ground truth, where applicable, to derive performance indicators is the basis for monitoring activities.

- **Problem and incident management:** Regulated companies should plan for problem and incident management processes that allow for targeted follow-up actions to mitigate deficiencies in data or in the model, or the integration into the AI-enabled computerized system. Product and process understanding and data understanding are important to allow for sufficient responsiveness to unforeseen and/or unexpected events; see Appendix P3.

- **Change management:** Change management should be integrated into quality monitoring. In addition to remediation and/or enhancement activities, change management allows for continual improvement, which may lead to new model versions or an extension of models based on additional data becoming available; see Appendix P3.

- **Management review of process performance:** Management oversight is required to establish and execute process governance. As such, information from ongoing monitoring, as well as supporting processes (i.e., risk management, data governance, change management), is suitably aggregated and prepared. This allows for informed decisions leveraging critical thinking throughout the life cycle, determining long-term consequences on the AI-enabled system and the wider organization.

Management is responsible for establishing a collaborative environment across stakeholders; see Appendix S2.

Figure 12.2: Quality Management System Pillars and Their Relationship to AI-enabled Computerized Systems

13 Appendix M2 – Supplier Management

13.1 Introduction

Regulated companies may collaborate with suppliers for various purposes in the context of AI-enabled computerized systems. They should effectively manage their suppliers to build trusted relationships in order to achieve high-quality systems.

This appendix provides guidance on leveraging supplier involvement to ensure the effectiveness and safety of AI-enabled computerized systems, in addition to Chapters 6 and 7. Relevant considerations include:

- Assessment of supplier capabilities and their maturity
- Alignment on data management practices
- Determination of the suitability of AI methods in the regulated company's context of use
- Transfer of knowledge regarding the AI approach
- Leveraging supplier records and information for verification and testing activities
- Planning ongoing performance monitoring of supplied models
- Implementation of change management processes particularly the management of model versions
- Integration of incident and problem management aspects, including particularities of AI

Guidance provided here focuses on AI-specific aspects and should be used in parallel with *ISPE GAMP 5 (Second Edition)* [2] and *ISPE GAMP® Good Practice Guide: Enabling Innovation – Critical Thinking, Agile, IT Service Management* [85].

13.2 Concept Phase

13.2.1 Supplier Assessment

The extent and level of the supplier assessment should be based on the supplier's risk, which is determined by two main components:

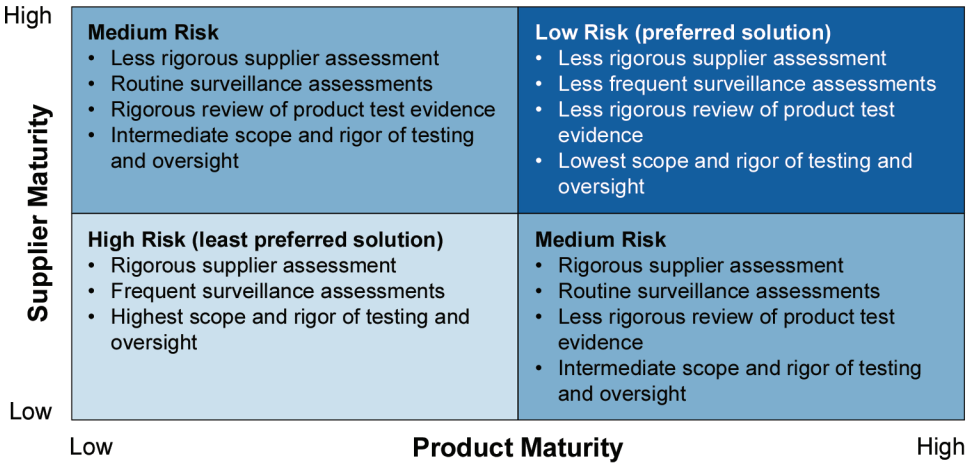
- The impact of the supplier on the intended use of the AI-enabled computerized system. The criticality of a supplier can change over time.
- The extent of delivered items or services, including data sets, a pre-trained model, and a software product with AI sub-systems.

The practical implementation of a supplier assessment is described in detail in *ISPE GAMP 5 (Second Edition)* [2], with additional information present in standards such as ISO 19011 [86] and other sources [87].

When evaluating potential suppliers of data sets, models, or AI-enabled software products, regulated companies should assess suppliers' regulatory understanding, compliance considerations, and their approach to quality.

As a result, they may categorize suppliers according to their maturity and the maturity of their products (see Figure 13.1), which can inform risk management processes.

Figure 13.1: Supplier and its Product Maturity (adapted [88])



13.2.1.1 Supplier's Understanding of the Regulatory Landscape

While a supplier is not directly regulated by GxP, the regulated company should ensure compatibility of their activities with their obligations. In addition, general regulatory requirements apply. To determine potential gaps and initiate activities to support supplier's regulatory understanding, the regulated company should consider the following aspects as part of their supplier assessment:

- **GxP regulations:** An understanding of GxP regulations such as 21 CFR Part 11 (US) [89] and EU GMP [90] is vital for successful collaboration. While they do not include AI-specific aspects at the time of publication, regulated companies and suppliers should be aware of the evolving regulatory landscape in the life sciences context concerned with AI.

- **Privacy regulations:** Depending on the model's context of use, privacy regulations need to be considered, including implications of individuals exercising their right of data to be removed, and its consequences on data and models.

- **International and regional laws:** Depending on the geographical scope of operations, AI-enabled computerized systems may need to comply with multiple regulatory standards across different regions. For example, deploying AI-enabled medical devices in the EU requires the organization to comply with the requirements of the EU AI Act [24]. As similar regulatory frameworks may evolve, the supplier should monitor changes in applicable regulations and demonstrate compliance if directly affected.

- **Ethical standards:** Beyond legal requirements, ethical standards play a critical role when using AI-enabled computerized systems. Regulated companies should ensure the compatibility of their supplier's activities or their software products with trustworthy AI principles; see Appendix M9.

13.2.1.2 Assessment Criteria for Supplier's Quality Capabilities

Regulated companies should consider the following AI-specific assessment criteria when assessing suppliers:

- **GxP impact:** Awareness of potential impact of the supplier's services on patient safety, product quality, and data integrity

- **AI regulations and standards:** Awareness of emerging AI-specific regulations and standards

- **Ethical AI development and use:** Consideration of ethical guidelines in AI development

- **Secure development practices:** Practices enabling AI security throughout the development life cycle
- **Model and data management and governance:** Practices and controls in handling data and models
- **Assurance of privacy:** Practices to protect training data, model parameters, and user data
- **Third-party risk management:** Practices in managing sub-supplier relationships, including data, models, or software products
- **Model integrity and protection:** Practices to safeguard models from tampering, unauthorized access, and theft
- **Bias detection and mitigation:** Processes for identifying and mitigating biases in AI models
- **Transparency and explainability:** Documentation on model architecture, decision-making processes, and facilitation of interpretability
- **Ongoing monitoring and updates:** Coverage and capabilities of monitoring model performance, detecting anomalies, and possibilities of providing timely updates
- **Incident response and recovery:** Handling of incidents specific to failures of AI-enabled computerized systems

13.2.2 Collaboration During Concept Phase

As described in Appendix P1, building understanding occurs early in the concept phase. This results in collaborative formats between regulated companies and suppliers, including ideation and feasibility assessments.

13.3 Project Phase

13.3.1 Understanding of Data and Models

During the project phase, regulated companies need to establish an understanding of the data and models used as part of the planned AI-enabled computerized system. Relevant aspects include:

- Data understanding
- Shortcomings of data quality
- Suitability of models and their limitations
- Implications of design decisions, such as fine-tuning
- Benchmarks from comparable settings

They should ensure integration and synchronization with stakeholders on the supplier's side.

13.3.2 Model Design Space and Model Requirements Specifications

When considering software products, the model design depends on the supplier's product's capabilities or its integration into software products. Navigating the choices needs careful consideration; for instance, fine-tuning a model may improve the model's performance, but add complexity to change and configuration management in operation. Regulated companies should make design decisions based on a thorough process and product understanding and the objectives of models in the context of use, following a risk-based approach.

Capabilities provided by the supplier regarding model performance indicators inform model requirements specifications. For example, performance indicators derived from the confusion matrix may be standard in a classification case, although tailored KPIs capturing risks specific to the regulated company's context of use may require access to more granular data.

13.3.3 Iterative Experimentation

When executing iterative experimentation to derive a suitable model and model configuration, capturing information and records on model engineering and configuration, as well as performance evaluation, in each step support the model development process, building a rationale for the model selected.

Considerations regarding software products that admit iterative experimentation include:

- The flexibility in data and model engineering capabilities
- The degree of automation in keeping records and information during iterative experimentation
- The possibility to compare performance of models across iterations

13.3.4 Model Testing

The performance verification of chosen model candidates and their configurations should be based on independent, unseen data that is representative of the regulated company's context of use.

Suppliers may provide information on additional tests performed on the use of models in their products in comparable settings, so that the regulated company can compare and challenge the performance determined during model testing activities specific to their context of use. The supplier can further support these activities by sharing experience in improving the performance of models and highlighting potential model shortcomings relevant to the risk control strategy.

13.3.5 Testing of the AI-Enabled Computerized System

Regulated companies can take advantage of the experience that suppliers have gained from similar installations, as well as supporting the assessment of sufficient coverage of real-world scenarios based on their experience with other clients; for further considerations on testing of AI-enabled computerized systems, see Appendix P2.

13.4 Operation Phase

Regulated companies should consider the following aspects when operating AI-enabled computerized systems:

- **Change management:** Regulated companies should understand how suppliers manage changes. They should carefully evaluate potential model updates provided by the supplier, particularly in cases where fine-tuning was applied, to avoid unintended consequences and potential misses of performance expectations when switching to a new model. They may align with suppliers to redeploy a previous model version when necessary.
- **Communication and resolution of incidents:** Regulated companies may expect transparent and timely communication from suppliers, so that regulated companies can react accordingly, e.g., tightening control during operation.

  If the issue has been corrected and a new version is provided, regulated companies may expect information on implications of the correction and further actions required.

- **Support for incident management and incident response plans:** Regulated companies should ensure that they have adequate means to analyze incident root causes, given the suppliers' products' capabilities.

- **End-user feedback:** Often, AI-enabled computerized systems include means for end users to provide feedback given model results. An overview or summarized reports help regulated companies use such feedback for decision-making on potential changes or their risk control strategy.

13.5 Retirement Phase

Post operation, regulated companies should consider the retention of data and models in line with guidance provided in Appendix P4. The regulated company should specify retention in agreements, including provisions on:

- Access to retained data and models

- The linkage of data and models to demonstrate traceability on the model version applied to what model input to derive what model output

- Availability of functionality, such as XAI methods, to reconstruct the situation for *ex post* assessment, if applicable

If the supplier and the regulated company are involved in decommissioning the system, performing a controlled shutdown, or disposing of the system, the retirement plan should be reviewed and mutually agreed upon. This plan should clearly outline the roles and responsibilities for activities under the supplier's responsibility.

If retirement includes supplier offboarding, a plan for system disposal or ownership transfer should be established, executed, and tracked.

14 Appendix M3 – Science-Based QRM

14.1 Introduction

This appendix provides further detail on the QRM process introduced in Chapters 2 and 5.

As stated in *ISPE GAMP 5 (Second Edition)* [2]:

> *"QRM is a systematic process for the assessment, control, communication, and review of risks to patient safety, product quality, and data integrity, based on a framework consistent with ICH Q9. …It is used:*
>
> - *To identity risks and to remove or reduce them to an acceptable level*
> - *As part of a scalable approach that enables regulated companies to select the appropriate life cycle activities for a specific system"*

Organizations should have established risk assessment methods and tools; see Sections 14.8 and 14.9. This Guide suggests a risk management approach for AI-enabled computerized systems that should be used in addition to the processes and techniques described in *ISPE GAMP 5 (Second Edition)* [2] as part of an overall QRM process. Critical thinking should be integrated into the QRM process; see Appendix M4.

14.2 Guidelines and Regulations

Several regulatory frameworks and guidance provide GxP regulated companies with varied approaches and methods for risk management. A combination of ICH Q9(R1) [31] and ICH Q10 Pharmaceutical Quality System (PQS) [35] provides a high-level, non-prescriptive approach to QRM.

For additional guidance and recommendations on wider aspects of QRM, consider the following resources:

- *ISPE GAMP 5 (Second Edition)* [2]
- *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* – CSA Appendix [1]
- ISO/IEC 23894 [8], which provides a comprehensive framework for the application of risk management methodologies for AI systems
- ISO 31000: Risk Management Standard [91] offers a procedure, structure, and set of guidelines for risk management
- AAMI TIR34971: Application of ISO 14971 to machine learning in artificial intelligence—Guide [92] provides guidance on applying risk management principles to AI-enabled devices

The following aspects illustrate specific considerations on QRM; they are meant as examples, not as a complete list:

- **High patient risk versus regulatory impact:** EMA distinguishes in their reflection paper between patient risk and regulatory risk: *"This paper uses the term 'high patient risk' for systems affecting patient safety, while the term 'high regulatory impact' is used for cases where impact on regulatory decision-making is substantial."* [43] In this appendix, patient risks are primarily addressed, while robust risk management practices described here may also support adherence to expectations regarding regulatory decision-making.

- **Medical devices:** Considering AI-enabled medical devices, organizations should follow guidance such as the hazard analysis principles in ISO 14971 [7] or the FDA guidance on the content of premarket submissions for software contained in medical devices [53]; further guidance is presented in Appendix S6.

- **Generative AI:** The NIST AI 600-1 framework [93] emphasizes that generative AI can exacerbate existing AI risks while introducing unique risks. These risks can vary across multiple dimensions, including the stage of the AI life cycle, scope (individual model, application, or ecosystem level), source of risk, and time scale. Organizations may tailor their approach to measuring Generative AI risks based on characteristics such as model architecture, training mechanisms, data type(s) used, levels of model access, and application context; see Appendix S3.

  As an additional example, the Federal Office of Information Security (Germany) guidance [94] categorizes risks according to:

  - Proper use: The Generative AI model is used as intended, although it may provide misleading or erroneous output.

  - Misuse: The Generative AI model is used in its original function, although being exploited in a way that may *"lead to a general erosion of trust in (media) content."*

  - Attacks: The Generative AI model, or its AI-enabled computerized system, falls victim to a malicious actor ("attacker") that *"deliberately and intentionally attempts to disrupt the function of an IT system or to gain access to it."*

Specific risks based on these categories should be used to derive adequate control strategies. Considerations should be interpreted in the context of the GxP computerized system's use case, i.e., its implications for patient safety, product quality, and data integrity. A collection of resources is available from the Federal Office for Information Security [95].

For additional information on cyber security specific aspects of AI-enabled computerized systems, see Appendix S5.

14.3 Benefits

Applying QRM to critical aspects of a computerized system in a controlled and justified manner leads to benefits such as:

- Identifying and managing risks to patient safety, product quality, and data integrity
- Scaling life cycle activities and associated records according to system impact and risk
- Justifying use of supplier documentation
- Better understanding of risks and controls
- Highlighting areas where detailed information is needed
- Improving business process understanding
- Supporting regulatory expectations

Specific to the use of AI, the benefits expected are:

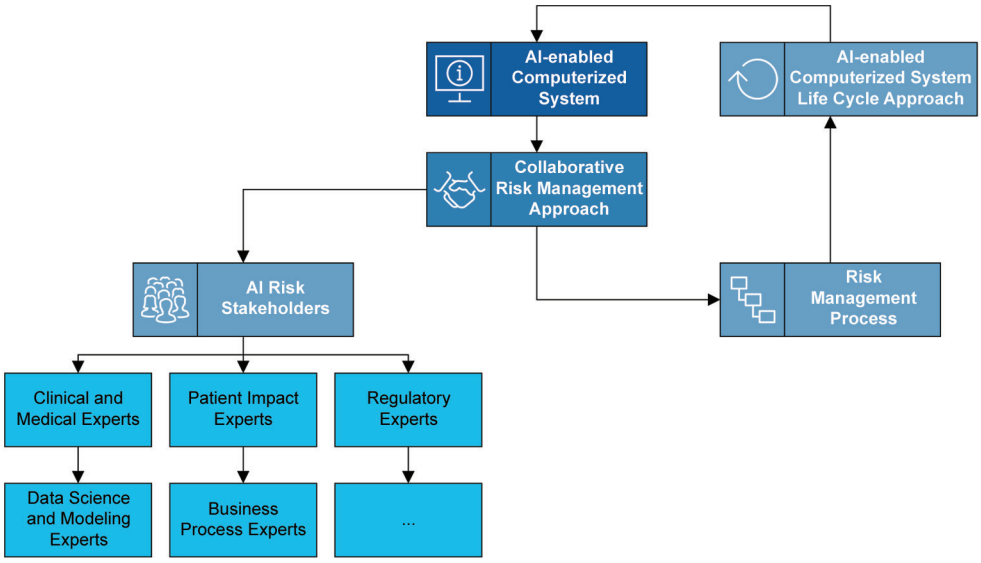
- Choosing adequate models and methods that ensure safety and effectiveness of AI-enabled computerized systems; these include the use of XAI methods to support human-AI interaction, see Appendix S4
- Adhering to trustworthy AI principles, see Appendix M9
- Improving data understanding in establishing the relation between data's fitness for purpose and risk; see Appendices M6 and M7
- Enabling innovation with the use of new technologies while maintaining patient safety, product quality, and data integrity

14.4 Roles and Responsibilities

QRM is the responsibility of the business process owner, which may be delegated to project members.
QRM is more effective if a diverse set of stakeholders are included,[17] ideally skilled at critical thinking and equipped with sufficient AI literacy.

The relation between stakeholders, risk management, and the AI-enabled computerized system life cycle is illustrated in Figure 14.1. Roles and responsibilities are summarized in Table 14.1, along with key activities.

Figure 14.1: Collaborative Risk Management Approach



[17] The NIST AI 600-1 framework [93] emphasizes the importance of involving diverse stakeholders in risk management processes to achieve a holistic view. It suggests engaging *"interdisciplinary teams that reflect a wide range of capabilities, competencies, demographic groups, domain expertise, educational backgrounds, lived experiences, professions, and skills across the enterprise to inform and conduct risk measurement and management functions."*

Table 14.1: Risk Management Activities by Roles

Stakeholder	Expertise	Responsibilities	Key Activities
Process Owner	Domain knowledge, business requirements	• Define the intended use • Provide input on data selection and model evaluation	Hazard identification, risk assessment, risk control definition
Business Analyst	Domain knowledge, business requirements, technology expertise	• Ensure consistent process understanding • Gather and structure requirements of the system • Ensure suitability of the model development approach	• Translation of business requirements to technical team members • Support the process owner for decision-making and planning
Data Owner (as an additional role or in conjunction with additional responsibilities of the process owner)	Data understanding, data governance principles	Advise on identifying suitable data sources, ensure adherence to data governance principles	Data identification, provision of information on origin and shortcomings in available data
System Owner and Sub-system Owner	• System overview, infrastructure and integration, performance • Infrastructure, security, and maintenance	• Ensure adequate infrastructure and security controls, monitor system performance • Determine requirements on integration capabilities of the system • Advise on technical feasibility of models based on available infrastructure	• Deployment and release, change and configuration management • Assessment of data quality in operation, human interaction • Performance monitoring
Project Manager	Time and resource management	Project management	Project risk identification, resource allocation, timelines, and business expectations
Data Scientist, Data Engineer	Data analysis, feature engineering, model selection	• Assess data quality and representativeness • Select appropriate model architecture	Data split, model design, model engineering, model evaluation
ML and AI Engineer, Operations	Model development, integration, and deployment	• Implement and train models • Integrate into production environment	Model engineering, deployment, and release
Software Engineer	Software development, integration, and validation	• Develop software components • Ensure proper integration of AI model	Model integration, deployment, and release
Tester	Software quality assurance	• Verify software components • Ensure integration of AI model according to specifications	Verification of fulfillment of requirements and model performance
Quality Unit	Regulatory requirements, validation, data integrity, and change management	Ensure compliance with applicable regulations and internal standards	Risk assessment, risk control definition, risk review

Table 14.1: Risk Management Activities by Roles (continued)

Stakeholder	Expertise	Responsibilities	Key Activities
Training Department	Training of personnel	Design training program in support of model use case	Creation of training activities for human/model interaction
Senior Management and Other Leadership Roles	Strategic direction, resource allocation	• Provide overall guidance and support • Ensure alignment with organizational goals and risk tolerance	Risk review, resource allocation for risk mitigation
External Stakeholders	Regulatory expectations, patient perspective	• Provide input on regulatory expectations and patient needs • Participate in risk communication	Evaluation of the appropriateness of risk assessments and risk-based decisions
Data Steward	Oversight perspective of data, data integrity	Coordination and control of overarching aspects regarding data, data integrity, and governance	• Risk assessment, control, and review with a holistic view • Critical thinking to identify gaps
ML Architect	System architecture and integration of AI sub-systems into the computerized system	• Translate business requirements into system architecture and design strategy • Consider the internal mechanics and inherent risks of a model within the system	Risk assessment, control, and review from an integration and technical point of view on the computerized system
IT Infrastructure and IT Services Units	Infrastructure- and IT-related requirements (including robustness and security), operation practices	• Focus on the boundary of the AI-enabled system and its perimeter, including interfaces to other systems • Evaluate long-term performance and scalability • Support design of guidance for operation	• Risk assessment, control, and review from a technical IT operation point of view • Coordination with other IT functions
Cybersecurity Specialists	Security-related requirements and best practices focused on IT systems	• Support the safe implementation and use of AI-enabled systems • Safeguard IP and corporate and personal data	• Risk assessment, control, and review from a cyber security point of view, including assessment of adversarial attack vectors on the AI-enabled system • Periodic vulnerability assessments
Ethics, Legal, and Data Protection Personnel	• Human-related requirements • Handling personally identifiable information, environmental, social, and governance aspects including diversity, equity, etc.	• Raise awareness for ethical aspects of AI implementation and responsible AI principles • Protect and safeguard personal data (if applicable) • Review representativeness and potential bias in data and models	• Risk assessment, control, and review from an ethics and legal point of view • Coordination with cyber security specialists and the Quality Unit

14.5 Scalability of the Process

ISPE GAMP 5 (Second Edition) states: *"the five-step risk management process may be scaled according to risk, complexity, and novelty of individual system, with each step of the process building upon the previous output."* [2]

14.6 Applying Risk Management Based on the Business Process

Per *ISPE GAMP 5 (Second Edition)* [2], effective application of QRM requires *"a thorough understanding of the business process supported..., including the potential impact on patient safety, product quality, and data integrity. Aspects to consider include:"*

- What are the hazards?
- What is the harm?
- What is the probability of failure?
- What is the detectability of failure?
- How will the risk be managed?

"Zero risk is usually an unattainable goal, and there are diminishing returns as it is approached." Regulated companies should also consider that risks may have multiple potential impacts. [2]

Given the above questions, the following aspects are worth noting:

- The use of AI may introduce new types of hazards rooted in the relevance of data and use of complex models; thus, identification of hazards should be expanded by characteristics specific to AI. See Section 14.7.3.
- The probability of failure is connected to the performance of models integrated in AI-enabled computerized systems; expectations on their performance needs to be defined in the model's context of use; see Appendix P2.
- When using complex models, the detectability of failures may be impaired by a lack of comprehensiveness on how the model has derived its model output from model input, limiting interpretability. See Appendix S4.
- While general risk management methods apply, including controls that reduce probability or increase detectability, specific methods may be required; see Section 14.7.3.

14.7 Risk Management Throughout the System Life Cycle

> *"Appropriate risk-management processes should be followed throughout the life cycle to manage identified risks and to determine the rigor and extent of the activities required at each phase of the life cycle."* [2]

In the following subsections, topics specific to the use of AI in computerized systems are provided, concerned with typical risk-based decision-making processes.

14.7.1 Initial Risk Assessment

"An initial risk assessment should be conducted at (or before) the beginning of the project phase." [2] As described in Appendix P1, regulated companies should determine the system impact including a decision whether the system is GxP regulated, gather information on functions with impact and capture possible hazards as well as initial ideas for controls.

"The assessment should be based on an understanding of business processes and business risk assessments." [2]. Additional assessments may be needed based on the complexity of the process and the complexity and nature of the anticipated system. Data understanding in conjunction with process and product understanding should inform such assessments. In addition, experience from previous AI projects should be considered, effectively leveraging knowledge management.

14.7.2 Risk-Based Decisions During Planning

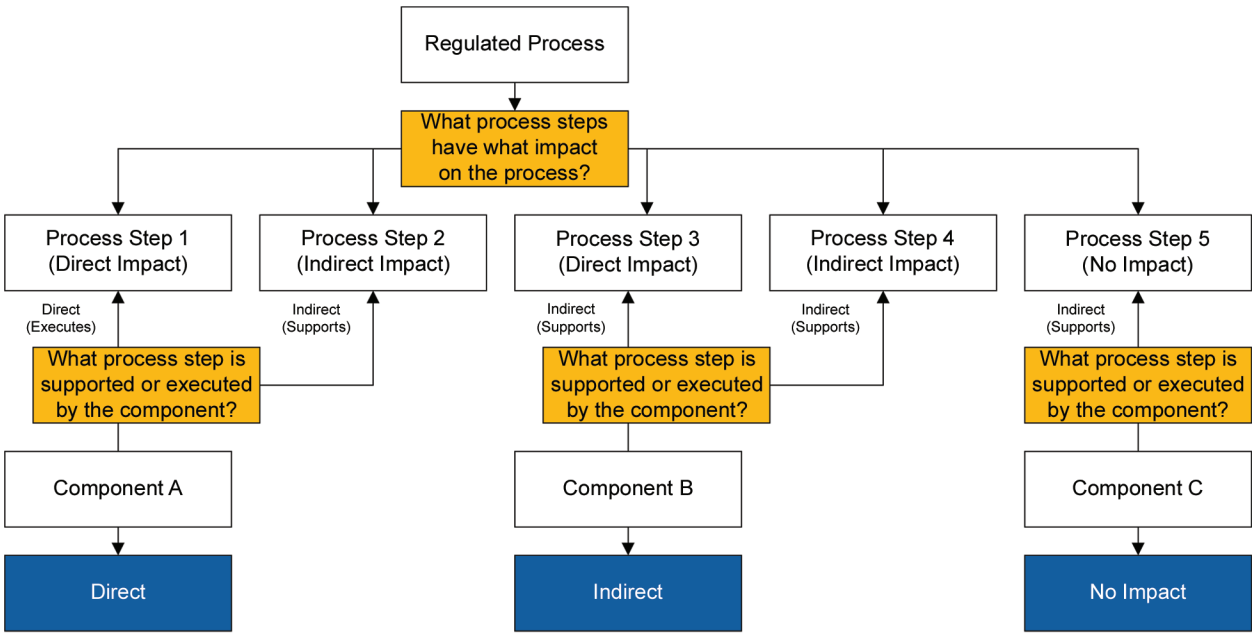
Risk-based decisions during planning inform subsequent process steps. Per *ISPE GAMP 5 (Second Edition)* [2], key aspects include:

- Need for, and rigor of, supplier assessments
- Use of supplier assessments for planning, for example, determining the involvement of the supplier
- Determining activities, deliverables, and responsibilities
- Need for further risk assessments, when they are required in the life cycle, and the methods used

Planning should also include a mapping of process steps, including determination of their impact[18] and individual components and their role in supporting one or many process steps. Such a mapping informs subsequent steps, such as the functional risk assessment, where AI-specific considerations need to be made. A schematic mapping is illustrated in Figure 14.2.

The use of AI may influence how the system is compartmentalized in individual components and their respective role in relation to process steps.

Figure 14.2: Schematic Mapping between the Regulated Process, Process Steps, and Components



[18] The draft FDA guidance "Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products" [55] considers a) impact of the system on reliability of results from nonclinical or clinical study and b) AI is used to produce information or data intended to support regulatory decision-making.

14.7.3 Functional Risk Assessment

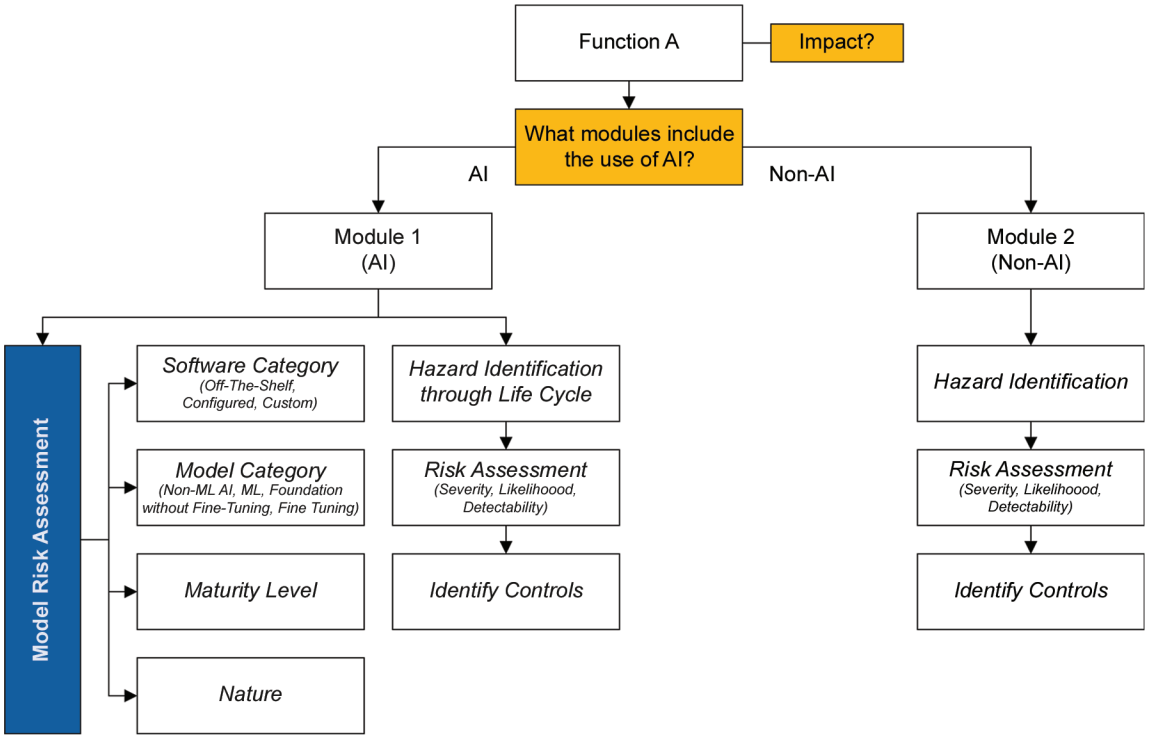
"Where these are required, functional risk assessments should be used to identify and manage risks to patient safety, product quality, and data integrity that arise from failure of the function under consideration." [2] It also influences the rigor and focus of testing activities with the goal of focusing on those areas that carry higher risk, thus applying critical thinking.

Functional risk assessments should consider AI-specific aspects, aspects related to the integration of AI within the AI-enabled computerized system, and functions not linked to the use of AI; see Figure 14.3. Components should be divided into functions related to AI and those not related to AI:

- Functions that do not use AI can be evaluated using a typical functional risk assessment
- Risk assessment of AI functions should consider the following nuances that will inform the choice of controls and the rigor of testing activities:
  - A model risk assessment should include model categories (non-ML, ML, foundation without fine-tuning or fine-tuned models; see Appendix M11), the maturity level (see Appendix M10), and the nature of the AI sub-system (complexity and novelty).
  - Individual hazards should be identified and assessed, considering risks that may be observed (Section 14.7), their severity, likelihood, and detectability, and controls identified (Section 14.10).

The functional risk assessment should consider the context of use and whether additional information is used alongside the model output to address the intended use.

Figure 14.3: Functional Risk Assessment for AI and Non-AI Functions and Modules



Model Influence per FDA Draft Guidance

In their draft Guidance for Industry and Other Interested Parties: Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products [55], the FDA defines model influence as *"the contribution of the evidence derived from the AI model relative to other contributing evidence used to inform the question of interest."* The model influence relates to the AI functional risk assessment as follows: the model influence is a result of risk-based design decisions, determining which influence the model development approach is entrusted with, while control measures, such as human involvement, may lead to additional decision-making influencing elements. Therefore, such measures limit the influence of the model, where a degree of model evolvement should be chosen commensurate with the use case risk and the experience of the organization applying the model development approach, for instance.

Risk assessments can focus on the functional risks of the computerized system (including non-AI-enabled computerized system components), as well as the AI sub-system, specifically the risks associated with the AI model as outlined in the FDA draft Guidance. Both considerations should inform the appropriate system design and control strategies.

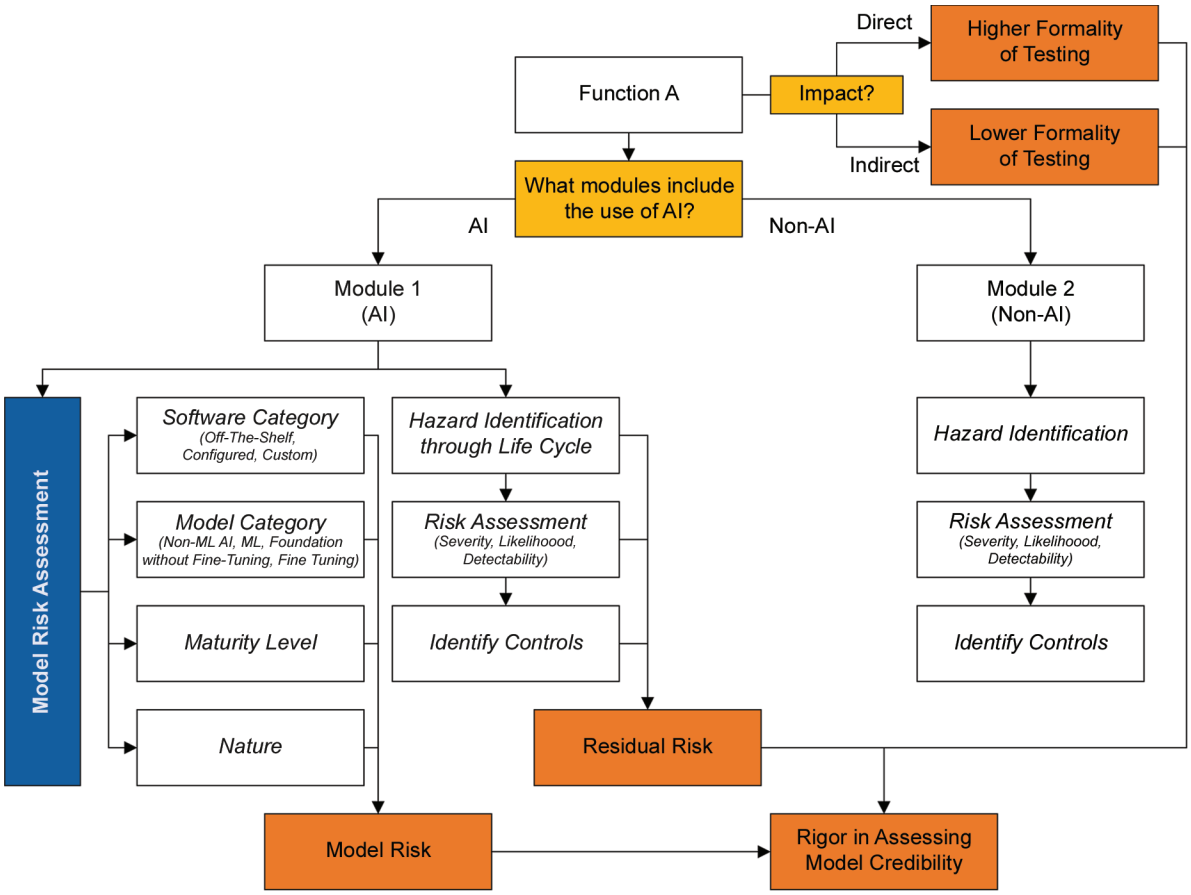
14.7.4 Risk-Based Decisions During Test Planning

> *"Testing is often performed at several levels depending on the risk, complexity, and novelty of the system. …The results of functional risk assessments should influence the extent and rigor of verification."* [2]

Regulated companies should consider the results of the functional risk assessment specific to AI (Figure 14.4):

- The model risk assessment can be aggregated to a model risk, which informs data expectations on testing activities, for example, the properties of test data set and acceptance testing scenarios (see Appendix P2)
- Identification of hazards and their risk assessment and controls yield residual risks that inform the rigor applied when assessing model credibility (see FDA draft guidance [55]) and performing testing activities, in conjunction with the model risk and the impact of the AI sub-system on the process

Figure 14.4: Use of Functional Risk Assessment Results for Test Planning



14.7.5 Risk-Based Decisions During Planning of Operational Activities

"Operational activities should be selected and scaled according to the nature, risk, and complexity of the system in question," with options provided in *ISPE GAMP 5 (Second Edition)* [2].

Specific to AI-enabled computerized systems is the planning of an adequate monitoring strategy that considers hazards stemming from the involvement of data and models and the human-AI interaction; see Appendix P3.

14.7.6 Functional Risk Assessments in Change Control

> *"Change management should provide a dependable mechanism for prompt implementation of technically sound improvements following the approach to specification, design, and verification."* [2]

There are changes that stem from the possibility to improve models by the use of new data, as well as changes required from incident and problem management or CAPA management processes; see Appendix P3.

14.7.7 Risk-Based Decisions When Planning System Retirement

> *"Risk-based decisions are required when planning system retirement, [including the] approach to data and record retention, destruction, or migration, [and the] approach to verification."* [2]

Additionally, the maintenance of the traceability, and reproducibility (as applicable) of model input, model versions, and model output to allow for *ex post* assessments need to be considered for AI-enabled computerized systems; see Appendix P4.

14.8 Hazard Identification Method

Decisions made throughout model development may impact model performance in operation. Regulated companies should systematically identify possible risks throughout the life cycle of AI-enabled computerized systems, and activities in relation to their AI sub-systems.

Leveraging the life cycle model presented in Chapter 3, nine hazard groups help to structure risk identification (see Figure 14.5):

1. **Data quality of the case data set:** Shortcomings in the fitness for purpose of the case data set may negatively affect model performance in the operation phase; see Appendix P2.

2. **Data split:** If the data split yields a test data set that is not representative of data expected during operation, a performance drop may occur in the operation phase; see Appendix P2.

3. **Model design:** An inferior model design (e.g., selecting a suboptimal model configuration) may yield a less effective model for the context of use; this may lead to non-acceptable testing results, or, when requirements are passed, a higher-than-necessary risk to patient safety, product quality, and data integrity. It is a component of critical thinking and scalable life cycle activities to determine an appropriately sized model design space; see Appendix P2.

4. **Iterative experimentation:** The search for an optimum set of features, hyperparameters, or instructions to the model may be prematurely stopped, posing a higher-than-necessary risk to patient safety, product quality, or data integrity when the potential for improvement is present. As in model design (Item 3), decisions on continuation of iterative experimentation constitute an example of scalable life cycle activities; see Appendix P2.

5. **Testing the AI-enabled computerized system:** testing serves as safeguard before operation. Bias or errors in this performance assessment may yield an acceptance decision that does not align with the regulated company's established risk tolerance; see Appendix P2.

6. **Release:** Choosing the wrong AI sub-system or model version for deployment and release poses risks to patient safety, product quality, and data integrity in the operation phase, since the deployed model may exhibit inferior model performance compared to accepted thresholds; see Appendix P2.

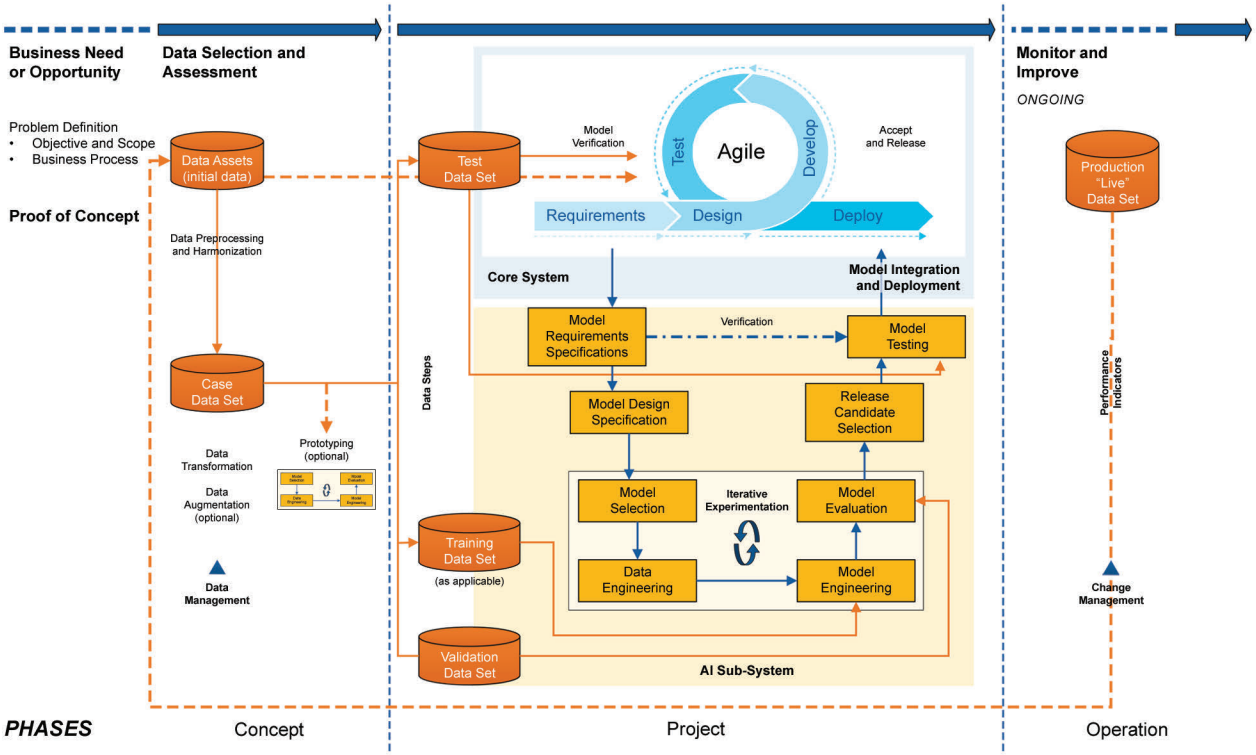
7. **Data quality in operation:** If data quality does not meet expectations compared to the basis upon which the evidence testing the model was generated, a loss in performance can be expected; see Appendix P3.

8. **Human interaction and monitoring:** The effectiveness of risk mitigation measures and human oversight, as well as the performance of the human-AI team (if appliable) depend on the design of monitoring and human interfaces and supporting processes such as training; see Appendices P3 and S4.

9. **Supplied models and software products:** involvement of suppliers may introduce risks, such as inadvertent changes to models, or the low performance of models for the specific context of use of the regulated company; see Appendix M2.

Risks related to cybersecurity are relevant throughout all phases of the life cycle. See Appendix S5.

Figure 14.5: Overview of AI-specific Hazard Groups throughout Life Cycles



14.9 Risk Assessment Methods

> *"Risk management aims to establish controls such that the combination of severity, probability of occurrence, and detectability of failures is reduced to an acceptable level."* [2]

As also noted in Steimers and Schneider 2022 [96] ,*"established risk mitigation measures in software development are only partially suitable for applications in [AI-enabled computerized systems], which only create new sources of risk. Risk management for systems using AI [should] therefore be adapted to the new problems."*

The following overview of techniques and tools is intended to support an informed decision on the best-fitted approach in a particular use context:

- **Failure Mode and Effects Analysis (FMEA):** A methodical process that identifies possible failure modes, assesses how they could affect the product's performance, and suggests countermeasures for any malfunctions. FMEA can be used to evaluate the influence of various failure modes on the performance and outputs of AI systems by analyzing the model, data, and infrastructure components.

- **Risk Analysis and Mitigation Matrix (RAMM):** The RAMM model integrates individual hazards with process step risk ratings, extending the FMEA. It allows for an analysis of impacts of hazards on quality dimensions [32], see Appendix M1.

- **Software Hazard Analysis (SHA):** A technique for identifying and assessing possible risks connected to software-based systems. The process entails assessment of the software's architecture, design, and code to detect any risks and mechanisms of failure [97]. In the case of AI-enabled computerized systems, it should cover data management, system integration, model development and model integration. SHA considerations are an integral part of AI-enabled medical devices and can be developed by following the collective methodologies of ISO 13485 [98], IEC 62304 [6], and ISO 14971 [7], see Appendix S6.

- **Hazard Analysis and Crucial Control Points (HACCP):** Although HACCP has historically been applied to food safety, its ideas can also be applied to AI-enabled computerized systems to help identify crucial points where hazards can arise during the AI life cycle and to put measures in place to prevent or mitigate them.

- **Hazard Operability Analysis (HAZOP):** An organized and methodical review of a proposed or current process to find and assess issues that pose a danger to people or property. HAZOP can be applied to AI-enabled computerized systems to examine the process of developing and implementing AI, spotting possible deviations and spotting deviations and their possible implications.

- **Preliminary Hazard Analysis (PHA):** A semiquantitative examination to find possible risks and unintentional events that could cause patient harm. PHA can be applied early in the development process to detect possible risks related to the planned use and design of AI.

- **Complementary statistical instruments:** Risk assessments in AI-enabled computerized systems can be supported and informed by a variety of statistical techniques, including control charts, Pareto charts, histograms, and process capability analysis. These tools are especially useful for examining performance data and spotting trends or abnormalities.

- **What-if analysis:** An organized brainstorming method that is helpful for investigating edge cases or unanticipated events and can be utilized to find possible risks and their effects in AI-enabled computerized systems.

- **Fault tree analysis:** This technique of deductive reasoning is used to identify the underlying causes of a certain undesirable event. It entails drawing a logic diagram showing the possible combinations of errors that result in the undesirable outcome. Fault tree analysis is a useful tool for analyzing model failure, data quality problems, and system malfunctions in AI-enabled computerized systems.

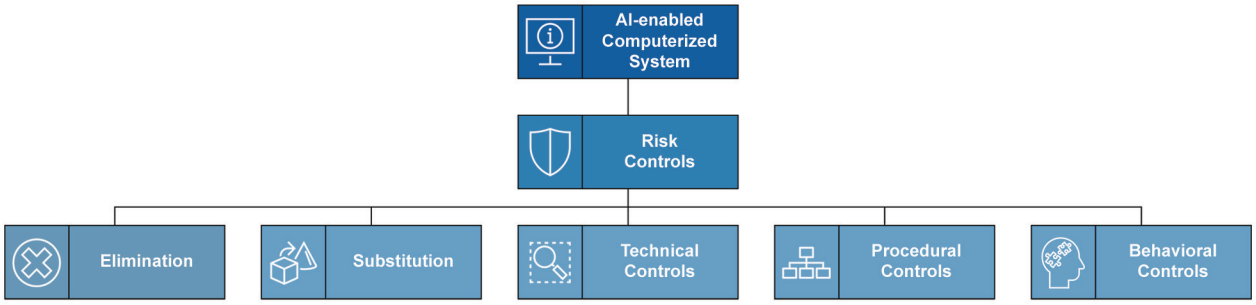
The selected risk assessment method should reflect the complexity of the AI-enabled computerized system and facilitate risk review, thus supporting continual improvement of systems.

14.10 Selection and Use of Controls

Based on the completed risk assessment, controls are identified to mitigate risks; "the organization shall determine the risks and opportunities according to the domain and application context of an AI system; the intended use; the internal and external context" [5].

Various risk control strategies can be employed to mitigate risks associated with the use of AI, in line with *ISPE GAMP 5 (Second Edition)* [2]: elimination, substitution, technical controls, procedural controls, and behavioral controls (see Figure 14.6).

Figure 14.6: Overview of Risk Controls



Collaboration between a wide set of stakeholders and multiple iterations is considered good practice to derive effective risk controls; experience from past use of AI in comparable settings should be leveraged.

14.10.1 Elimination

Risk elimination can be reached by individual risk reduction measures identified in the functional risk assessment and applying suitable control measures. In addition, the choice of a lower AI maturity level may be considered to reduce risks; see Appendix M10.

14.10.2 Substitution

Substitution is a technique that involves changing elements that carry too high a risk with an alternative option that carries a lower risk. For instance, a complex model type may be substituted by a different model type that is easier to interpret by end users; see Appendix S4.

14.10.3 Technical Controls

Technical controls refer the implemented and supportive functionality enabling risk reduction. Elements of technical controls specific to AI-enabled computerized systems include modularization and isolation, automation, and the use of model performance metrics, triggers, and alerts (see Appendix P2), as well as controls on cybersecurity aspects (see Appendix S5).

14.10.4 Procedural Controls

Procedural controls are structures, processes, and policies to ensure the AI-enabled computerized system is developed, validated, and operated in a controlled manner. These include establishing a thorough validation framework, data and model governance and management, knowledge management, as well as change management, risk management, and supplier management processes. Procedural controls are based on clear roles and responsibilities and agreed standards throughout the organization; see Appendices M2, M5, and M7.

14.10.5 Behavioral Controls

Behavioral controls rely on human oversight. While not all use cases allow for direct human oversight of every model result (e.g., AI-enabled visual inspection), human oversight is an important element by means of ongoing monitoring and initiation of interaction when required; see Appendices P3 and P4. XAI methods may support the effectiveness of end users' interaction with AI; see Appendix S4, Section 14.16, and Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles [33].

The design of behavioral controls should integrate insights from validation activities and be subject to ongoing monitoring during operation to ensure their effectiveness.

To maintain effective human oversight, training and education should raise awareness of AI limitations in general by means of AI literacy and regarding the specific context of use; see Appendix M5.

14.11 Residual Risk

"Residual risks after implementing control measures should be considered. … If the residual risk is above the threshold of acceptable risk, then appropriate further controls should be implemented and verified, and the impact on previously implemented risk control measures should also be considered." [2]

14.12 Scaling Life Cycle Activities

Risk considerations are the basis for scaling life cycle activities; see Appendix M11, and *ISPE GAMP 5 (Second Edition)* [2].

14.13 Risk Communication and Documentation

Considerations on risk communication and documentation of *ISPE GAMP 5 (Second Edition)* [2] apply; in particular:

- The output of the risk management process should be shared by the decision makers with other involved parties, including the quality unit (where necessary and appropriate)
- Communication should take place throughout the risk management process
- Communication does not necessarily take the form of a report
- Attention should be given when a risk or impact has changed
- Risk management output may inform change management processes

14.14 Risk Management for Outsourced Activities

See *ISPE GAMP 5 (Second Edition)* Section 11.5.9 [2] and *ISPE GAMP® Good Practice Guide: Enabling Innovation* Appendix M11 [85] for guidance on risk management for outsourced activities.

14.15 Risk Review

The risk review for AI-enabled computerized systems should verify the adequacy of risk controls. Insights from ongoing monitoring, periodic review of risks, refinement of risk assessments, and the control strategy are required to ensure a state of control. See Appendix P3.

The AI maturity, the experience of the organizations, the novelty of the AI-enabled computerized systems, and the level of risk can be used to determine an appropriate risk review cycle.

In addition to information gained from using the AI-enabled computerized system, insights from similar computerized systems should be included, thus cross-fertilizing understanding of risks and knowledge across the organization; see Appendix M5.

The risk review may result in changes to the AI-enabled computerized system or AI sub-system design, the chosen maturity level, or the choice of controls.

14.16 Risk Considerations when Applying XAI

Two perspectives on XAI methods are relevant in connection with risk management activities: while XAI methods can support risk mitigation strategies, they introduce their own complexities and risks.

The following implications of XAI **support** risk control strategies:

- **Facilitate user adoption and improve effectiveness of the human-AI-team:** End users, particularly those less experienced or literate in AI, may be reluctant to adopt AI. XAI methods may support adoption by demonstrating the linkage between model input and model output, allowing for meaningful interactions, thus facilitating effective human oversight while increasing human expertise and experience. Not all cases require direct oversight; see Appendix M10.

- **Support in RCA:** In the event of problems or incidents, XAI methods can help identify root causes or model flaws and inaccuracies, allowing for a targeted approach to model improvements, and provide a clearer view on model limitations; see Appendix P3.

- **Achieve robust and effective models:** XAI methods can support prototyping, design, and experimentation during model development, which can lead to targeted models. For example, XAI methods can help identify potential biases in data and models, producing effective models that adhere to trustworthy AI principles (see Appendix M9). In addition, XAI methods can help identify the strongest features and patterns, enabling model simplification, reducing models to essential drivers of performance, and avoiding potential overfitting; see Appendix P2.

- **Support decisions on model choices:** XAI methods support an understanding of the drivers behind model output, enabling SMEs to link the model with their product, data, and process understanding. This builds rationales on model choices, following a risk-based approach, and thus promoting quality and compliance with regulatory expectations; see Appendix P2.

The following **risks** should be considered when choosing XAI methods [99]:

- **Misguidance and target group inadequateness:** Even though XAI aims to help users draw the right conclusions from model outputs, XAI can still be affected by the limitations and inadequate quality of the data used. Furthermore, users may misinterpret the output.

- **Mismatch between the data used for inference by the model and for XAI:** The same data and state of model should be used for determining the actual model output and the result from XAI methods. Misalignments may cause confusion and provide misleading information to users, which can limit the effectiveness of XAI methods.

- **Lacking level of detail:** Explainability cannot reveal the full mechanics of the model, which would be as complex as the model itself. Therefore, a decision on the appropriate level of abstraction needs to be made to determine a level of information that suits user‘s needs and AI literacy. This level of detail should be carefully evaluated and tested to ensure its sufficiency.

- **Infrastructure burden:** Given potentially high requirements on computing power, XAI methods may be too resource-intensive or slow to provide operational benefits in the context of use. This aspect should be carefully evaluated in infrastructure planning and when designing the AI-enabled computerized system and adjacent interfaces, specifically when handling limited infrastructure, e.g., in the context of medical devices (see Appendix S6).

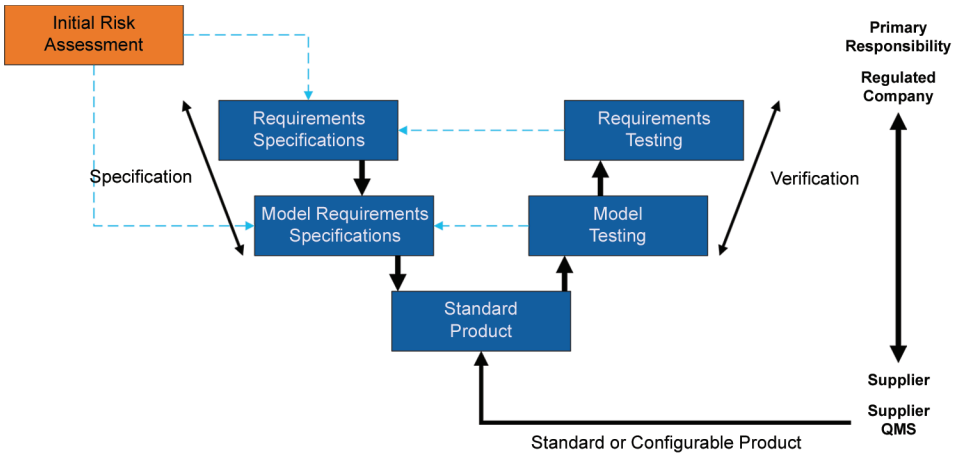
14.17 Examples

This section presents examples of the application of risk management meant to be indicative and not definitive.

14.17.1 Example 1: Standard Product

For a standard product, risk management activities may be bundled in a single assessment as shown in Figure 14.7. Depending on the complexity and nature of the system and the experience of the regulated company, further assessment may be necessary.

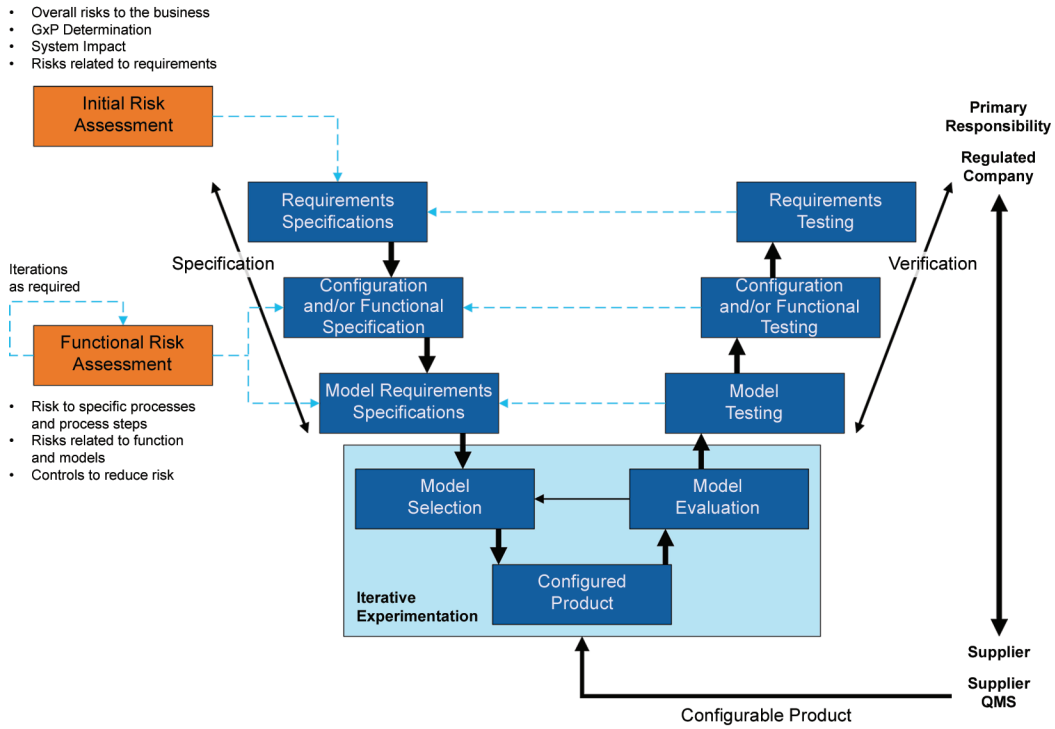
Figure 14.7: Risk-Based Approach for a Standard Product (GAMP Software Category 3)



14.17.2 Example 2: Configurable Product

For a typical configurable product, an initial risk assessment and functional risk assessments (and iterations thereof) covering configurations and/or functional specifications of non-AI components as well as model requirement specifications may be performed; see Figure 14.8. More extensive or less complex approaches may be chosen, depending on the complexity and nature of the system, and the experience of the regulated company.

Figure 14.8: Risk-Based Approach for a Configurable Product (GAMP Software Category 4)



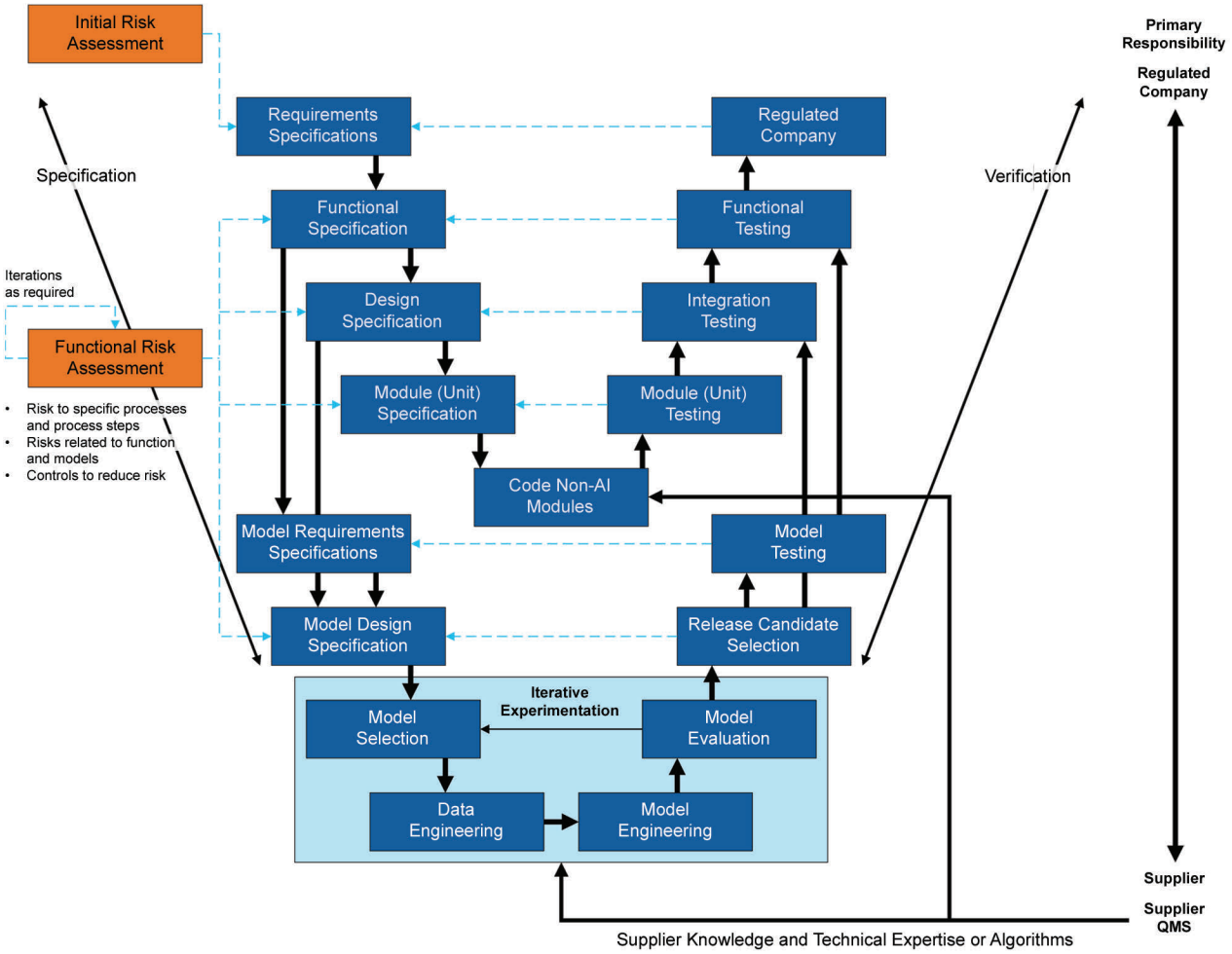
14.17.3 Example 3: Custom Product

For a custom product, an initial risk assessment and multiple functional risk assessments (and iterations thereof) may be performed. See Figure 14.9. Functional risk assessments in this example apply to:

- Functional risk assessment based on functional specifications
- Functional risk assessment based on design specifications
- Functional risk assessment based on model (unit) specifications
- Functional risk assessment based on model requirements specifications
- Functional risk assessment based on model design specification

More extensive or less complex approaches may be chosen, depending on the complexity and nature of the system and the experience of the regulated company.

Figure 14.9: Risk-Based Approach for a Custom Product (GAMP Software Category 5)



15 Appendix M4 – Critical Thinking

15.1 Introduction

Product and process understanding is key for a holistic assessment of risks. In addition, data understanding is highly relevant given the role of data throughout the AI-enabled computerized system life cycle. Building such an understanding should include contributions and views from a diverse set of stakeholders.

Product and process understanding as well as data understanding are expanded and deepened throughout the life cycle as new insights are generated on the interplay of data and models, and from aspects of human and AI interaction. These insights serve as a key input for risk management activities, as described in Appendix M3.

Leveraging products and process understanding, data understanding, and risk management activities gives rise to the use of critical thinking and scalable life cycle activities.

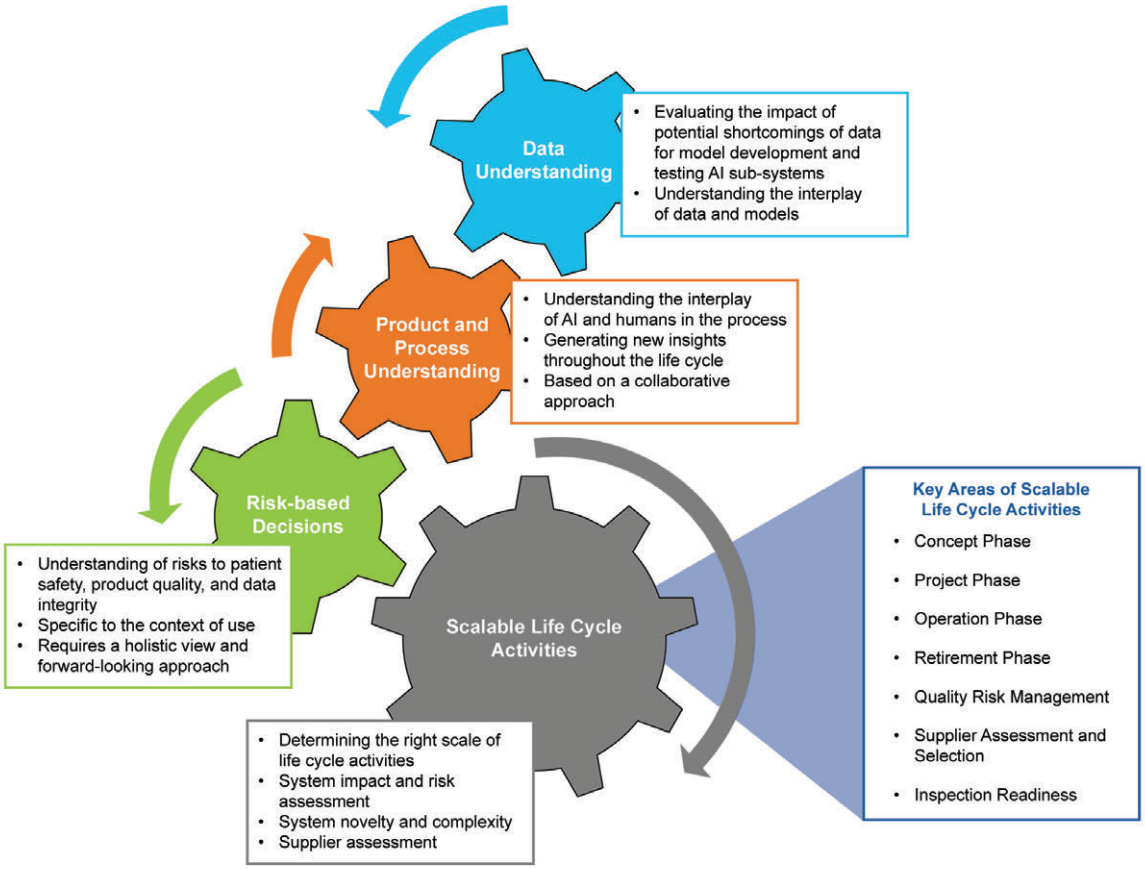
Life cycle activities should be scaled according to [2]:

- The system impact on patient safety, product quality, and data integrity based on the risk assessment
- The system complexity and novelty
- The outcomes of the supplier assessment (if applicable)

Scaling life cycle activities may change as more is learned about the AI-enabled computerized system.

Given this background, this appendix provides guidance on typical drivers for choosing the appropriate scale of such activities to derive decisions throughout the life cycle (see Figure 15.1). It covers considerations throughout the concept, project, operation, and retirement phases, QRM, supplier assessment and selection, and implications for inspection readiness.

Figure 15.1: Scalable Life Cycle Activities Overview



Scalable Life Cycle Activities and Critical Thinking

ISPE GAMP 5 (Second Edition) Appendix M12 [2], elaborates on critical thinking to promote informed decision-making and good judgement in determining where and how to apply and scale quality and validation activities for computerized systems. The critical thinking concept is based on the observation that overly rigid processes and "tick-the-box" methods may lead to situations where resources and effort are wasted, or risks are even overlooked, instead of focusing on more valuable and essential quality activities. Overcoming such rigid approaches contributes to efficiently scaling life cycle activities to maximize impact, while simultaneously meeting expectations on patient safety, product quality, and data integrity.

Example: Illustration and Considerations on Varying Risks Depending on the Context of Use

Risk to patient safety, product quality, and data integrity vary considerably depending on the context of use. For instance, as the EMA reflection paper on the use of AI in the medicinal product life cycle [43] points out, *"applications of AI/ML in relation to indication or posology are regarded as a high patient risk [case],"* while *"applications within pharmacovigilance may allow a more flexible approach to AI/ML modelling and deployment."* However, these risks to patient safety, product quality, and data integrity should be evaluated in the context of a particular use case, going beyond a broad categorization (see Appendix M3).

Additional resources on critical thinking are *ISPE GAMP 5 (Second Edition)* [2], *ISPE GAMP Guide: Records and Data Integrity* [15], and *ISPE GAMP Good Practice Guide: Enabling Innovation* [85].

15.2 Concept Phase Considerations

AI is a dynamic field, offering a range of innovative approaches. However, careful consideration should be applied to support use cases where the use of AI provides a promising proposition to raise the quality of operations and serve business objectives. Therefore, critical thinking should include organizational aspects, such as culture, knowledge, and AI literacy, as foundations for the successful application of AI.

Once a suitable use case is identified, a typical element of the concept phase is to perform a feasibility assessment and create a prototype. Regulated companies should apply critical thinking to arrive at a reasonable extrapolation of these preliminary results concerning realistic expectations on the potential impact, the generalizability of the approach, and its compatibility within the AI-enabled computerized system and end users.

Experience from lessons learned in previous AI projects may help achieve an objective evaluation.

See Appendix P1 for additional details.

15.3 Project Phase Considerations

The project phase is concerned with identifying an adequate model that meets defined requirements and leads to successful release of the AI-enabled computerized system.

Managing numerous choices regarding data modeling and model development aspects requires critical thinking to develop a comprehensive series of experiments throughout the model development process. A well-founded rationale for the selected model relies on process understanding and QRM activities, as well as an understanding of suitable controls. In particular, the depth of experiments conducted during iterative experimentation and the selection of relevant data are examples of scalable life cycle activities.

When risks to patient safety, product quality, or data integrity are deemed high, additional experimentation may be conducted to meet higher thresholds on KPIs, compared to situations where the system impact is deemed lower.

The design of suitable control measures can also be considered a scalable activity based on information gathered during the project phase, establishing processes for adjustments during system operation.

Lastly, testing activities should be planned based on an understanding of risks, including considerations on the quality of test data sets and data used for testing the AI-enabled computerized system. In addition, the rigor of testing activities needs to be determined.

See Appendices P2 and M3 for additional details.

15.4 Operation Phase Considerations

Scalable life cycle activities relate to numerous activities during operation to maintain patient safety, product quality, and data integrity.

The level of detail in education and training materials, as well as the cycle to which education and training plans are executed, should be considered scalable life cycle activities, with choices commensurate with the users' capabilities and per risk assessment.

To this end, performance indicators and quality control techniques may provide empirical evidence on the effectiveness of human oversight, substantiating training effectiveness or indicating needed changes.

AI sub-systems, specifically when applying ML, offer the chance to learn from new data generated during operation, either via model version management with verification steps or fully automated in dynamic systems. At the same time, certain risks arise, including model drift or changing environmental factors. To balance the various drivers, change management requires critical thinking when seeking to improve model performance, aiming for reducing risks and supporting business objectives.

The cycle upon which such decisions are made and the evolution of model version review can be considered a scalable life cycle activity, with decisions based on the empirical evidence collected in the project phase and insights generated during operation. More specifically, the choice of performance indicators and the frequency of periodic review should be determined following a risk-based approach and critical thinking.

See Appendix P3 for additional details.

15.5 Retirement Phase Considerations

During the retirement phase, determining whether data and models can be used for other purposes should be completed. For instance, data may be used as part of an expanded training set or for further evaluation of other models, while models may be employed in other AI-enabled computerized systems or reactivated.

Such decisions require careful planning and consideration of legal obligations regarding records retention and data privacy. Considerations on an organizational level include data and model governance and management.

See Appendices M7 and P4 for additional details.

15.6 QRM Considerations

As stated in *ISPE GAMP 5 (Second Edition)* Appendix M12 [2], critical thinking is an important part of QRM activities. AI-enabled computerized systems vary in the way they pose risks to patients; see Chapter 5 and Appendix M3 in this ISPE AI Guide. Choosing control measures commensurate with the risks is integral to life cycle activities of AI-enabled computerized systems.

While QRM considerations generally guide decisions on scalable life cycle activities, decisions on the rigor and depth of QRM itself are required. For instance, functional risk assessments may be conducted on various levels or summarized in a fewer number of assessments.

A collaborative effort involving diverse stakeholders, providing a holistic perspective that augments each other's views, supports successful risk management and comprehensive decision-making, as does an open-minded culture that respects differing opinions.

An exchange of experience on QRM strategies and results is important to foster a common approach and best practices.

See Appendix M3 for additional details.

15.7 Supplier Assessment and Selection

Regulated companies need to consider differences in AI-enabled computerized systems compared to non-AI-enabled computerized systems during supplier assessment activities and the supplier selection process.

While building knowledge and experience with the use of AI in GxP areas, careful management of suppliers is required. This includes considerations on data management and model development practices.

Critical thinking may help decide whether to use AI. It can also aid in choosing between categories of models and supplier integration (e.g., AI-enabled software products off-the-shelf, the use of foundation models, or custom developed AI sub-systems).

Furthermore, critical thinking supports selecting suppliers of sufficient maturity and stability in the dynamically evolving context of AI.

See Chapter 6 and Appendix M2 for additional details.

15.8 Inspection Readiness

The mindset of always being ready for inspection has a positive impact on the culture and awareness in an organization in relation to quality and risk-based decisions. See *ISPE GAMP 5 (Second Edition)* Appendix M12, Section 20.3.11 Inspection Readiness [2]. Maintaining awareness of these factors assists stakeholders and functional areas in adequately allocating resources and efforts, thus rightsizing the rigor and scale of life cycle activities, including maintenance of the validated state.

Besides general factors such as insights from prior inspections or further exchanges with regulatory agencies and implications of risk management activities on achieving inspection readiness include:

- Overview of roles and responsibilities, including AI-specific roles or augmentation of roles in the context of AI.
- Trainings, according to the individuals' roles, including assurance of AI literacy.
- Understanding what areas and what level of detail should be covered by what role. For instance, while an end user is expected to demonstrate an understanding of data, model output, and XAI methods (as applicable), limitations, and their relevance of oversight and feedback, they may refer to SMEs for in-depth questions such as the rationale of the model design or insights from iterative experimentation relevant to the control strategy of the live system.
- Preparation of further information, for example, the selection and creation of test data sets, model testing, ongoing performance monitoring, and change management aspects including implications of the adaptiveness of AI sub-systems and rationales for decision made.
- Preparation of additional information and records, including results from model testing and ongoing performance monitoring.
- Information on the history of model versions and traceability of model input, models, and model output; see Appendix M7.

Similar to non-AI-enabled computerized systems, contractual agreements with suppliers are needed to allow access to information under specified circumstances (see *ISPE GAMP 5 (Second Edition)* [2]). In this context, the relevance of data and data management practices, and model development and change management with respect to models should be considered.

A valuable approach to support inspection readiness is to give those who may interact with inspectors the opportunity to practice. In addition to the conduct of formal internal audits, mock audits and walkthroughs provide a safe environment to gain experience, fostering good practices across the organization. The decision on the frequency and depth of mock or internal audits can be seen as a life cycle activity, and their insights can be used to scale other life cycle activities.

16 Appendix M5 – Knowledge Management and Building AI Literacy

16.1 Introduction

Management of information and knowledge is key to facilitating effective collaboration among stakeholders, allowing them to master complex challenges. Relevant aspects include:

- Facilitate understanding how an AI approach may support raising quality and achieving business objectives
- Maintain a state of control of the AI-enabled computerized system
- Allow informed, comprehensive, and effective decision-making
- Facilitate continual improvement, relying on lessons learned
- Establish a sufficient level of AI literacy

This appendix presents guidance on using and augmenting a knowledge management system to enable adoption of AI in line with the *ISPE GAMP 5 (Second Edition)* [2] and the *ISPE Good Practice Guide: Knowledge Management in the Pharmaceutical Industry* [38].

AI literacy is relevant for all stakeholders involved in AI-enabled computerized systems. Information is also provided to help build AI literacy across the organization. Aspects of a robust organizational culture are in the *ISPE Guide Series: Advancing Pharmaceutical Quality – Cultural Excellence* [100].

16.2 Knowledge Management

Several knowledge areas are relevant to AI-enabled computerized systems:

- **Data:** Contribution of data to business objectives or for improving quality
- **AI sub-system:** Selection of best-performing model designs and their integration in the process
- **Operating procedures:** Establishing governance and procedures to foster effective oversight and control
- **People:** Management of the human-AI interaction and cultural aspects to facilitate adoption
- **Equipment and hardware:** Requirements that arise from equipment and hardware, potentially implying boundaries on model choices
- **Operating environment:** Retaining and multiplication of knowledge among stakeholders

These aspects require thorough management of knowledge, including:

- **Handling knowledge throughout the AI-enabled computerized system life cycle:** A structured process should be established to capture, identify, review, analyze and disseminate knowledge to achieve knowledge availability [38]. For AI-enabled computerized systems and their AI sub-systems, key steps where knowledge is generated include:
  - The definition of the intended use, context of use and scope, as well as the identification of suitable data and methods, prototyping activities, and the assembly of the PoC information in the concept phase
  - Determination of model requirements specifications, iterative experimentation, identification and testing of the release candidate, AI-enabled system acceptance testing, and determination of the fitness of intended use in the project phase
  - Education and training, ongoing monitoring, and incident and problem management in the operation phase
- **Skill building for individuals involved in AI-enabled systems:** Established roles may be augmented, and new roles may emerge. This not only requires dedicated skill building so that individuals can effectively perform their activities, but also a collaborative approach to leverage deep expertise, spanning technical, business, and quality areas. Further information on typical roles and responsibilities is in Chapter 6.
- **Stakeholder resilience and multiplication of knowledge:** Individuals may build deep know-how in performing their tasks, including product and process understanding, and data understanding. However, people change positions or focus, which puts knowledge at risk. Therefore, information on key activities and insights should be maintained, while knowledge and insights should be shared. This not only reduces the risk of knowledge drain but also allows individuals to better connect with decisions made in managing the AI-enabled computerized system and its AI sub-systems, and support them in their activities.
- **Knowledge as a means of risk management:** Knowledge is foundational to comprehensive risk management activities and risk-based decision-making. As more knowledge is gained throughout the life cycle, it informs risk assessments and the design of control strategies. Efforts should be taken to utilize insights from previous risk management activities for comparable use cases across the organization, promoting a consistent approach and facilitating learnings in line with the *ISPE Good Practice Guide: Knowledge Management in the Pharmaceutical Industry* [38]. This requires openness and a willingness to share experiences, hence necessitates an open and trusted organizational environment to speak freely. Further information on risk management activities is provided in Appendix M3.
- **Data and model governance and management:** Knowledge should not only be seen in the context of the AI sub-system or its AI-enabled computerized system but also in its further value for the organization when leveraged for transfer, additional development, and adoption of AI. Therefore, organizations should introduce dedicated roles and responsibilities that support dissemination of knowledge about data and models. Data and models should be linked to stakeholders that can support deepening the understanding of their potential and limitations; see Appendix M7.
- **Supplier and regulated company collaboration:** Both regulated companies and suppliers typically hold knowledge that only in combination leads to effective and efficient development and use of AI. The transfer of knowledge needs to be managed to equip suppliers with an understanding of the intended use and context of use, and detailed requirements from a regulatory perspective, while regulated companies should obtain sufficient knowledge about data and models as part of the AI sub-system, and in how far they support the regulated process. Only based on trust relationships will knowledge be shared effectively; see Chapters 6 and 7 and Appendix M2.

16.3 Building AI Literacy

AI literacy includes foundational understanding of how AI methods function, the considerations required to activate such innovation effectively, and their implications within the context of use, which is included specific to regulation such as the EU AI Act [24]: *"'AI literacy' means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause."*

In addition, the EU AI Act [24] states the following requirement: *"Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used."* It is essential in the ideation, development, and use of AI-enabled computerized systems:

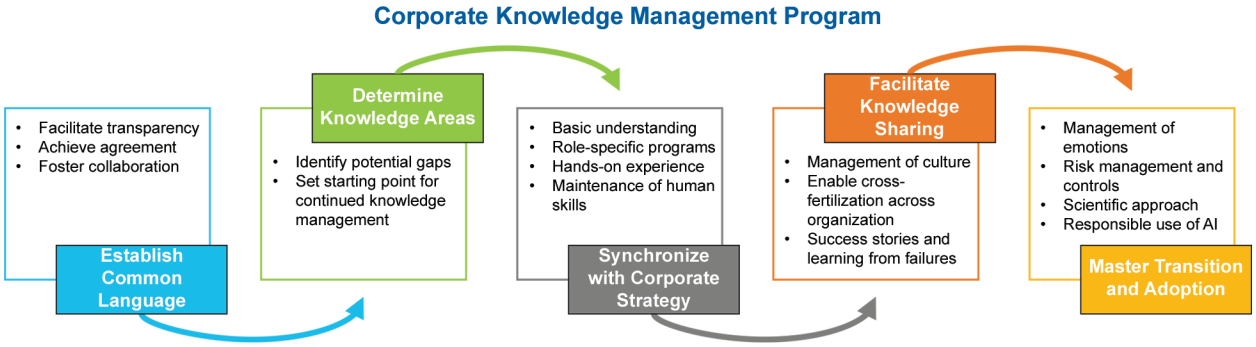
- **Ideation:** To contribute meaningful ideas for AI use cases, a basic understanding of AI is required to connect its capabilities with the current possibilities. Ideation typically occurs during the concept phase.

- **Design, development, and implementation:** Further understanding of AI is required to effectively navigate the development process, while drawing the right conclusions from generated evidence. Design, development, and implementation typically occur during the project phase.

- **Use:** To effectively perform their oversight role, users and operators need to be aware of potential weaknesses and limitations; also, they need to understand their crucial role in supporting further improvement of models in providing meaningful feedback. Use typically occurs during the operation phase.

Establishing AI literacy begins at the organizational level and transitions to an individual level. Figure 16.1 is an overview of the recommended program to build AI literacy.

Figure 16.1: Example Program to Build AI Literacy



Regulated companies should implement the following steps to advance AI literacy, thus establishing foundations for the successful adoption of AI:

- **Definition of a common language regarding AI:** A common language is the basis for effective knowledge management and collaboration. It is required to create transparency of the status quo as well as to facilitate a change to build skills and knowledge in this new area.

  A common language should be acceptable to all stakeholders, aligned and linked to the organization's structures, processes, roles, and technology. Additionally, key concepts should be established considering AI in computerized systems and agreed upon by all stakeholders.

- **Assessment of the current state of AI literacy and identification of gaps with the target state:** Establish a clear and transparent view on the regulated company's capabilities and processes to identify potential gaps in knowledge required for AI adoption. This will form a basis for the continued use of knowledge management when introducing AI-enabled computerized systems.

- **Synchronization of knowledge management activities with other strategic objectives in the organization:** Based on the status, a target vision and a roadmap for skill building and retention needs to be formed. In this regard, AI should not serve its own purpose but should practically support processes in GxP regulated and non-GxP regulated areas. Various approaches may be considered to form a solid base for the adoption of AI-enabled computerized systems:

  - A **basic understanding** of key AI concepts should be established, e.g., by fusing methodological background along with practical case studies. Such a basic understanding facilitates collaboration, but also ideation on suitable use cases. At the same time, it can mitigate both exaggerated skepticism and enthusiasm, thus facilitating a scientific, solution-oriented, problem-solving approach to AI that is fit for purpose and utilizes critical thinking.

  - **Role-specific programs** can help deepen the understanding of AI from different perspectives. For example, a data scientist may require technical skills in data management, data visualization, and model development, while a process owner needs to be able to identify and evaluate data risks as the basis for decision-making in the design of AI-enabled computerized systems.

  - Given the inherent **iterative nature** and the relevance of data for the implementation of AI-enabled computerized systems, **hands-on experience** is crucial. The combination of theoretical background and hands-on experience supports thorough risk assessment and comprehensive decision-making.

  - Lastly, as AI-enabled computerized systems are implemented, **practical experience in executing processes** run by these systems may be augmented or replaced. With regards to the latter, knowledge management activities should consider this diminishing of practical experience over time and how this can be remediated, if necessary.

  These aspects are compatible with typical knowledge management activities, integrating aspects relevant to AI with other expertise and skills. An open, flexible approach is recommended, combining strategic building of theoretical knowledge with careful consideration and management of opportunities to build hands-on knowledge.

- **Foster a culture of knowledge sharing:** Building a common language, regulated companies need to manage corporate cultural aspects to allow stakeholders to openly share their experiences in learning and applying new skills.

  Communication formats fitted to the organization should be established as part of the knowledge management plan to foster sharing experiences and exchanging best practices. Management needs to establish both a quality and knowledge sharing culture, carefully considering barriers to transparent communication. For example, communicating insights and lessons learned from a failed project may be difficult from an individual's point of view, although it is valuable for the organization.

  Setting clear and robust expectations on innovation projects, while allowing room to improve, is an important aspect of cultural management to facilitate effective knowledge sharing and a shared experience across the organization. See the *ISPE Guide Series: Advancing Pharmaceutical Quality – Cultural Excellence* [100] for details on corporate culture considerations.

- **Mastering adoption:** From both implementation and operation perspectives, transitioning from manual, automated, and hybrid systems to AI-enabled computerized systems will pose a challenge that should be addressed by organizational change management.

  Resistance to change is common, and many people perceive AI as a threat instead of seeing its adoption as potential option. Knowledge management, risk management, and control mechanisms outlined in this Guide can help overcome these challenges.

  In addition, a scientific approach, from providing a rationale on the use of AI to creating evidence via validation activities, can help manage fear while leveraging accepted and known concepts to bridge differences to non-AI-enabled computerized systems.

  For this reason, compatibility of concepts applied in risk management, change management, and other supporting processes is important to maintain efficiency, consistency, and cultural alignment. Monitoring adoption, including both failures and successes, can help manage the rate of change in alignment with the organization's culture and risk tolerance, ensuring the adherence to trustworthy AI principles.

17 Appendix M6 – Fit for Purpose Data and Data Quality

17.1 Introduction

This appendix covers the concept of fit for purpose data, which is data that meets expectations on reliability, relevance, representativeness, and abundance, as assessed for a model's context of use.
This appendix also covers data quality and specific data quality dimensions to support achieving reliable data, while guiding organizations in establishing data management standards.

Additional considerations on data governance may apply to GxP data; see *ISPE GAMP 5 (Second Edition)* [2] and *ISPE GAMP Guide: Records and Data Integrity* [15]. This appendix also expands on data governance aspects for non-GxP data that may be used for development purposes of an AI-enabled computerized system.

17.2 Fit for Purpose Data

Data is determined as fit for purpose, when the following four properties are met:[19]

- **Reliability:** Reliability describes the extent to which real-world relationships are adequately captured in the data.
- **Relevance:** Data should be relevant for the intended use. Regulated companies should carefully select and filter data to ensure the data used for model development and testing purposes is meaningful in its context of use, distinguishing it from unrelated data. See OECD Handbook for Internationally Comparative Education Statistics [101].
- **Representativeness:** Data should reflect the characteristics of the intended use and sufficiently mirror a range of real-world situations. In particular, samples should cover relevant population subsets.
- **Abundance:** The amount of data should be commensurate with the model's context of use and complexity. Generally, more data available leads to more meaningful results and model robustness. Data needs to be relevant to the purpose of the AI sub-system to achieve these goals, while covering the model input data space across relevant scenarios.

Note: Larger amounts of data may be more complex to process by humans and computers and are more costly to obtain. A compromise on the amount of data needed to build models that are fit for purpose may be required.

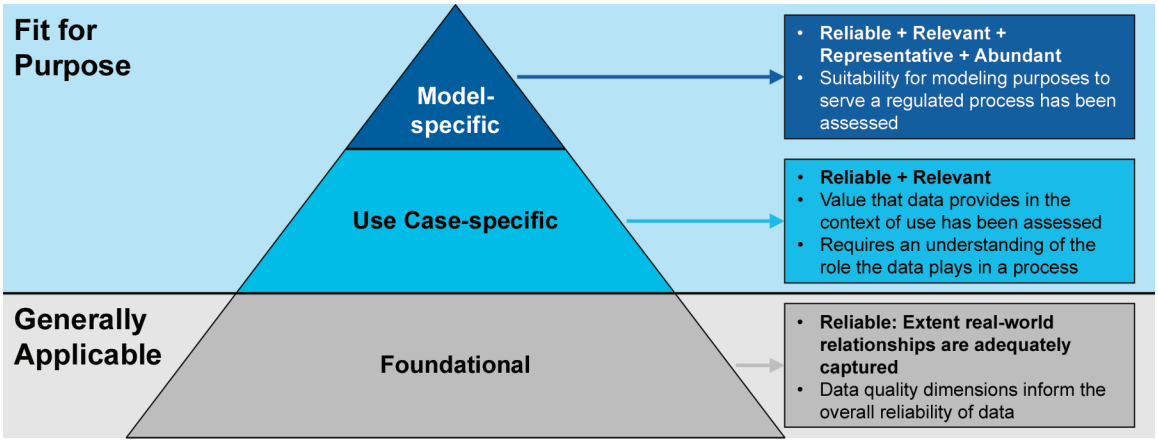
Reliability, relevance, representativeness, and abundance can be contextualized in how far they are specific to a model in the AI sub-system, as shown in Figure 17.1):

- **Foundational** considerations comprise various data quality dimensions as described in more detail in Section 17.3, leading to reliable data that appropriately captures real-world relationships
- **Use-case specific aspects** combine foundational considerations with an assessment of relevance, addressing the usefulness of data in the context of use of the model
- **Model-specific aspects** combine use-case specific and foundational aspects with representativeness and abundance, and qualifies data for use in a range of models.

[19] The concept of fit for purpose data is based on considerations of VDI [254].

Data that meets expectations on foundational, use-case specific, and model development specific levels is **fit for purpose.**

Figure 17.1: Hierarchy of Fit for Purpose Data



Regulated companies should link data's fitness for purpose with QRM processes; see Chapter 5 and Appendix M3. This allows for risk-based decision-making, choosing the appropriate effort to achieve data that adheres to the expectations of reliability, relevance, representativeness, and abundance.

Fit for Use and Fit for Purpose Data

The relevance of suitable data for development, evaluation, testing or use of AI models is emphasized in various regulatory (draft) guidance documents. For instance, EMA describes in their reflection paper [43] that *"efforts should be made to acquire a balanced and sufficiently large training dataset in relation to the intended context of use."* FDA draft guidance on the use of AI in the context of drugs and biological products [55] as well as medical devices [102] refer to the term of fit for purpose data. FDA [55] indicates that the terms fit for use and fit for purpose are sometimes used interchangeably, while this ISPE Guide uses "fitness for purpose" of such data sets. Furthermore, the FDA introduces key properties of data in their draft guidance [102] that relate to considerations in this section:

- **Abundance** is a desired property of the **quantity** of the data
- **Representativeness** is a desired property of **diversity** of the data
- **Reliability** is connected to the **quality** of the data

Relevance in the context of use is mentioned both in the FDA draft guidance [102] and this ISPE AI Guide.

17.3 Data Quality

As outlined in Section 17.2, regulated companies should assess whether data is reliable, irrespective of a specific use case, as a foundation to guiding their general data management activities.

Regulated companies can structure their assessment along data quality dimensions to determine the extent to which data appropriately represents real-world entities and relationships. While various concepts to describe data quality exist, two frameworks are commonly used across industries:[20]

- ISO/IEC 25012 [103] which lists inherent and system dependent dimensions
- OECD data quality framework [101]

On this basis, the following dimensions should be considered:

- Accuracy
- Completeness
- Consistency/coherence
- Credibility
- Timeliness/currentness
- Compliance
- Cost-efficiency/efficiency
- Precision
- Understandability/Interpretability
- Accessibility

Drawing the link between these data quality dimensions and the use of AI, Table 17.1 provides definitions and descriptions of the relevance of these dimensions. In this context, data quality dimensions as listed in Table 17.1 can be used to assess the data at hand and identify potential shortcomings as a baseline assessment (for instance in the concept phase).

[20] The discussion of suitable dimensions to assess data is not new; further publications include [104] and [105] from an academic point of view, while [106] collects 60 dimensions of data quality.

Table 17.1: Supporting Aspects and Relevance of Data Quality Dimensions in the Context of AI

Data Quality Dimension	Definitions (excerpts)	Supporting Aspects and Relevance in the Context of AI
Accuracy	OECD [101]: Accuracy is the degree to which the data correctly estimates or describes the quantities or characteristics that they are designed to measure. ISO [103]: The degree to which data has attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use.	Accuracy of data supports building robust and effective models, as well as deriving robust performance indicators during model evaluation and testing; however, achieving an appropriate level of accuracy requires considerations of model capabilities models to abstract from single data errors. In addition, labels in supervised training applications require high diligence, as they influence the development and the performance evaluation of the model.
Completeness	ISO [103]: The degree to which subject data is associated with an entity has values for all expected attributes and related entity instances in a specific context of use.	The complete data set should be considered for AI sub-system development, evaluation, and testing; any filters should be clearly justified with a rationale, e.g., limitations of the system's scope or statistical justification for excluding outlier values that pose risks such as instabilities or bias in the model.
Consistency/ Coherence	OECD [101]: Coherence reflects the degree to which the data are logically connected and mutually consistent. ISO [103]: The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities.	The consistency of data should be ensured when handling data from various sources and in various databases; the capacity of models to abstract from possible minor inconsistencies should be considered reflected, following a risk-based approach. This is particularly important during development, evaluation, and testing so that the representativeness of the data set is maintained when transitioning between development phases and system instances.
Credibility	OECD [101]: Credibility is the confidence that users place in data products based simply on their image of the data producer, i.e., the brand image. ISO [103]: The degree to which data has attributes that are regarded as true and believable by users in a specific context of use.	Data should not be blindly trusted; an assessment of internal data sources or external data providers should be performed with appropriate due diligence of the source or supplier. It is usually not sufficient to assess the suitability of such data superficially in a GxP context. Therefore, further analysis of the data should be performed to conclude its suitability in the context of use (see also Appendix P1 on EDA and assessment of plausibility of data).
Timeliness/ Currentness	OECD [101]: Timeliness reflects the length of time between data becoming available and the events or phenomena they describe. ISO [103]: The degree to which data has attributes that are of the right age in a specific context of use.	The reference point in time of input and output data allows for an assessment of model behavior throughout continued monitoring (e.g., to detect drift behavior). Examples of data where contemporaneous recording is relevant include labeling activities by users in a supervised learning context.
Compliance	ISO [103]: The degree to which data has attributes that adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use.	Noncompliance of data with standards, conventions, or regulations when used for model development purposes would render the derived model at regulatory risk. This should be considered particularly when sourcing third-party data.

Table 17.1: Supporting Aspects and Relevance of Data Quality Dimensions in the Context of AI (continued)

Data Quality Dimension	Definitions (excerpts)	Supporting Aspects and Relevance in the Context of AI
Cost-efficiency/ Efficiency	OECD [101]: Cost-efficiency measures the costs and provider burden relative to the output. ISO [103]: The degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.	Careful decisions should be made when acquiring data, considering project budget constraints while prioritizing the core objectives of patient safety, product quality, and data integrity. The incremental value of data (e.g., including populations not covered well by own data) should be assessed to make informed decisions about data acquisition. Similarly, efforts invested into own data pools (e.g., performing labeling exercises or employing methods to augment data) should follow a fair balance between expected benefits and efforts, commensurate with the risks introduced by shortcomings of data quality or coverage.
Precision	ISO [103]: The degree to which data has attributes that are exact or that provide discrimination in a specific context of use.	Understanding the precision of data is important, as lower precision may restrict the areas in which this data can be used. For instance, a lower precision or resolution in medical images may render the data nonvalid for certain types of models. These considerations are of particular interest when sourcing data from various sources, which may have various levels of precision.
Interpretability/ Understandability	OECD [101]: Interpretability reflects the ease with which users may understand and properly use and analyze the data. ISO [103]: The degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use.	AI-enabled computerized systems should be based on product and process understanding, and data understanding. Interpretable data helps to navigate design decisions in the development phase and to create models that match the understood real-world relationship between model input and desired output. However, the interpretability of data may be limited due to its mass and complexity, which should be reflected by risk management activities and controls.
Accessibility	OECD [101]: Accessibility reflects how readily data products can be located and accessed from within OECD data holdings.	Data needs to be available and retrievable for model engineering, evaluation, and testing (case data set) and during operation of the model (live data). Ensuring this may involve hardware considerations, such as the efficiency of data retrieval procedures or network speed, as required by the use case.

GxP Data and Data Integrity

In the context of GxP data, various principles have been established as regulatory expectations and are often already contained in existing regulations. These are known as ALCOA principles which define general requirements regarding the completeness, consistency, and accuracy of data [107]. ALCOA incorporates the principles of attributable, legible, contemporaneously recorded, original or a “true copy”, and accurate. The additional dimensions of complete, consistent, available, and enduring are often added to the ALCOA principles, known as ALCOA+ or ALCOA++ which includes the characteristic of traceability as mentioned in EMA guidance on computerized systems in clinical trials [108]. Guidance on GxP data principles is also provided in *ISPE GAMP Guide: Records and Data Integrity* [15].

While data for model development may be sourced from areas governed by data integrity (e.g., a manufacturing or clinical context), data integrity requirements are generally not expected for data used in model development, see [109]. For this reason, the more general data quality frameworks provided by ISO and OECD were used, many of which align with the data integrity principles, though imposing not the exact rigor on data life cycle management; see Appendix M7 regarding this topic.

18 Appendix M – Data and Model Governance and Management

18.1 Introduction

This appendix covers data and model governance and management aspects to support

- Traceability of technical artifacts, including data, models, software code, and performance indicators
- Linking the corporate perspective on data and models with operational aspects of a dedicated system
- Scaling of use cases and collaborative model development approaches

It considers the creation, modification, and use of data and models throughout life cycle phases, and their role in facilitating effective and safe AI-enabled computerized systems.

This appendix augments the approach described in *ISPE GAMP 5 (Second Edition)* [2] and *ISPE GAMP Guide: Records and Data Integrity* [15].

It primarily addresses regulated companies, who hold the accountability of implications of data and model governance and management; however, suppliers may use this guidance to meet the expectations of regulated companies.

The guidance provided in this appendix covers the following aspects of data and model governance and management:

- **Data and model governance framework:** A governance framework comprises organization and ownership, life cycle management of data and models, supporting processes, and nomenclature.
- **Data and model management practices:** Practices to create, curate, modify, tag, retrieve, and integrate data and models should be established, in line with context-specific guidance provided in Appendices P1, P2, P3, and P4. These practices ensure that data used during model development and testing is of high quality, see in Appendices M6 and P2.
- **Data operations and implications on models:** The interrelation between data operations (create, update, select, delete) and models needs to be considered. Creating new or making changes to existing data may have implications on models, while possible need for additional data capture or addressing data shortcomings triggered by model development activities can impact data.
- **Roles and responsibilities:** Data and model management have relevance on corporate level, Roles and responsibilities should be formed to address these implications and relations, augmenting considerations provided in Chapter 6.
- **Scaling AI Use Cases:** Data and model governance and management considerations apply to scaling use cases across the organization.
- **Data sharing and federated learning:** Data and model governance and management practices may facilitate collaborative settings in establishing AI-enabled computerized systems with third parties.

18.2 Data and Model Governance Framework

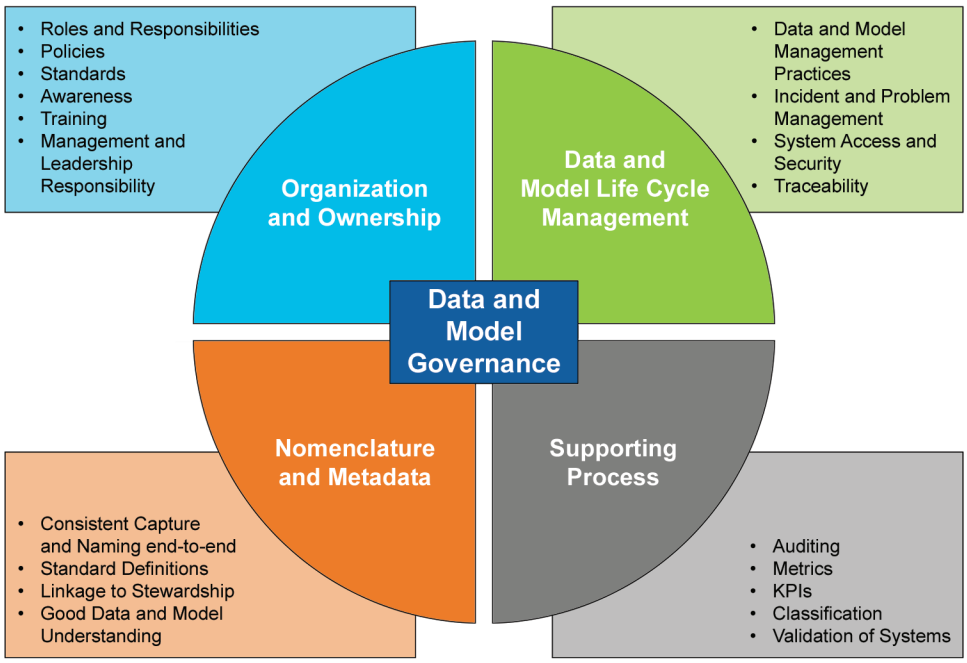
A data and model governance framework can be established by leveraging common data governance concepts, which are designed for GxP records and ensure that data, irrespective of how it is generated, is recorded, processed, and retained as a complete, consistent, and accurate record throughout the data life cycle [110]. Data and model governance also aims for legally compliant data use. On a more general level, considering both GxP and non-GxP data, various organizational and technical measures can be utilized in the context of AI-enabled systems [1, 15]:

- **Organization and data ownership:** Defining roles and responsibilities, policies, and standards, raising awareness, conducting training, and implementing leadership and management responsibility.
- **Data life cycle management:** Planning and executing QRM related to data, following robust data management principles based on adequate knowledge, establishing thorough data incident and problem management, and managing system access and security.
- **Supporting processes:** Auditing, establishing metrics and KPIs, and classifying data and validating systems that generate and store data.
- **Nomenclature:**[21] Framework to ensure that data is captured and named consistently from the beginning to the end of the business process, including coherent and reliable metadata. This includes master data management and standard definitions and should be linked to organizational aspects such as stewardship to maintain consistent and well-understood data.

These governance concepts should be expanded to model governance aspects, introducing life cycle management, change management and stewardship of models, as well as establishing rules to document the model origin, purpose, and applicable areas of use, see Figure 18.1.

Integration of various areas of expertise supports data and model governance, facilitating a collaborative approach.

Figure 18.1: Overview of Data and Model Governance Elements



[21] See *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* [1], where the concept of nomenclature was introduced.

18.3 Data Management

Robust data management practices ensure data quality, maintain cybersecurity, and allow for traceability of data usage, e.g., for development of models. This includes establishing data provenance for all data sets used. Therefore, standards should be established for:

- Data collection
- Handling of data sources
- Data access and provision
- Data transfer and communication between systems
- Data cleansing and transformation
- Syntactic transformation of data

A rationale should underpin the selection of specific data management practices and methods, integrating change management processes in line with risk assessment activities.

18.3.1 Types of Data

When considering data management activities, it is helpful to distinguish between three types of data, while considering challenges and typical data management activities:

- **Structured data** has a predefined format into which the data is categorized within a database, e.g., database contents, tables, and sensor data as time series. For example, data in relational databases can be represented by tables with rows (instances of data) and columns (fields of data), where fields can represent keys to other instances of data in related tables.

  The overall structure of a relational database is typically defined in a database schema, within which each field is given a name for identification and a data type. With a suitable level of detail in the database schema, important descriptive data (metadata), such as semantic meanings, units, and boundaries, are available.

  Implementing robust rule sets supports meeting data quality standards and maintaining accuracy of representations and descriptions of structured data.

- **Unstructured data** includes texts, photos, graphics, and other types of data.

  The semantic meaning of unstructured data should be identified and captured, enriching unstructured data with metadata or labels.

- **Polystructured data** refers to data that contains structured and unstructured parts whereby these parts are stored separately from each other. For instance, performance reasons regarding data retrieval or processing may favor polystructured data. Data storage systems that organize data in this way are referred to as polyglot data storage.

  Achieving a suitable level of data consistency when employing polystructured data management approaches can be a challenge, given the potential complexity of such data.

18.3.2 Data Collection and Handling of Data Sources

When utilizing data, it is usually necessary to collect, harmonize, and merge data from different data sources and transform it into a suitable format; examples include exploratory analysis and model engineering and evaluation or testing. Operational aspects of data collection and handling of data sources include the level of automation and ongoing monitoring.

A data inventory with contextualization and annotation of data with metadata allows for effective assessment of the suitability of a data set for a specific use case.

General strategy and architecture help improve data accessibility and usability across the organization, since data generated may be useful for purposes not yet anticipated at the time of data generation (e.g., a new model type). Data privacy considerations should be considered when handling sensitive data.

Regular monitoring of data can rely on visualization and statistical assessment, preferably supported by automation. This facilitates the identification of potential shortcomings in data quality. Regulated companies should establish follow-up activities for when shortcomings are observed or automated warnings are triggered, including inspection, decision-making, and potential remediation; see Section 18.8.

When establishing data management and monitoring practices, a variety of data sources may be of relevance; examples include files, databases, data lakes, document management systems, enterprise software, the internet and social media, external data sources, and data suppliers.

These data sources may be siloed or connected. Siloed systems refer to information systems, data, or processes that operate in isolation, often within specific departments or units of an organization, without seamless integration or communication with other systems. These “silos” create barriers that prevent the flow of information and thus may hinder use of data and collaboration across an organization. In contrast, connected or integrated systems allow for streamlining of processes and automation, e.g. via automated data transfer and harmonized capture of data.

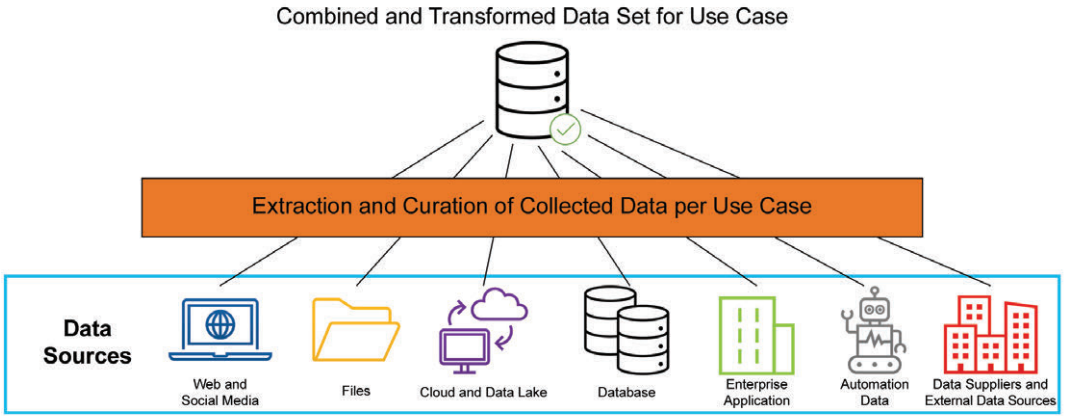
Suitable approaches for managing data vary in their levels of generalizability and adaptability, as well as upfront investment:

- **Direct data collection approaches** (low generalizability): By case collection and harmonization of data directly from data sources
- **Data collection strategy with global data storage** (medium generalizability): Centralized collection of data including metadata
- **Data collection strategy with global generic data objects** (higher generalizability): Conversion of data into generic data objects

Direct Data Collection Approaches

Direct data collection approaches employ integration, mapping, and matching of data to overcome schematic heterogeneity on a single case basis; see Figure 18.2. Structural heterogeneity may include data modeling choices in the respective data source, such as normalization, the choice of structured and unstructured elements, or different naming of tables and columns.

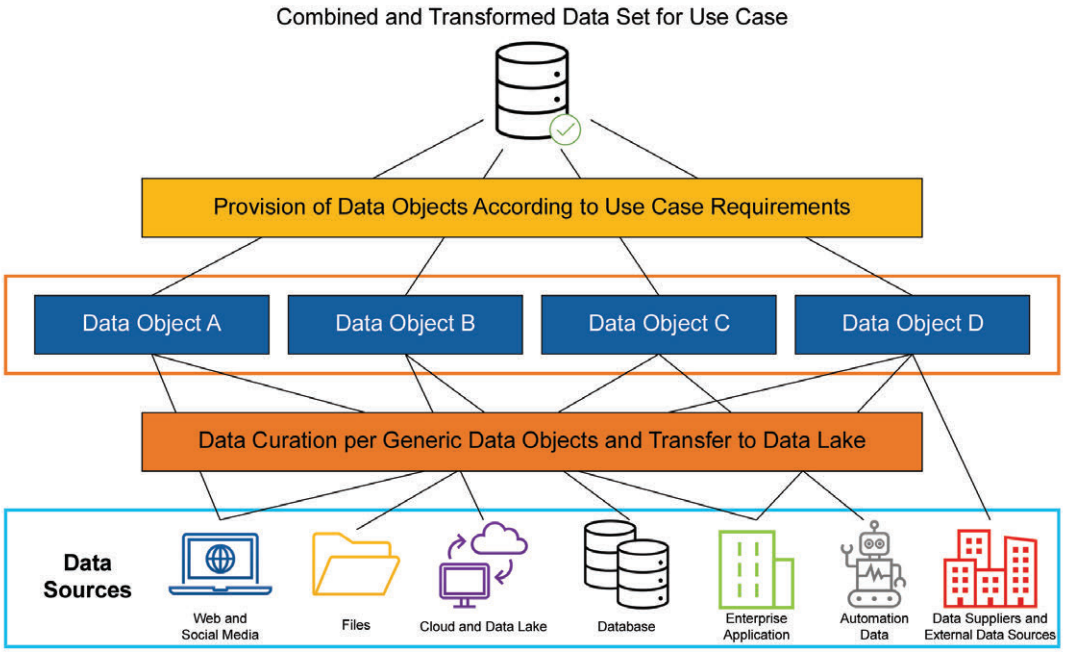
Figure 18.2: Illustration of Direct Data Collection Approaches



Data Collection Strategy with Global Data Storage

This data collection strategy is more flexible and yields higher reusability of data, as illustrated in Figure 18.3. Regulated companies establish a common ontology and collect data with attached metadata. This strategy supports teams to identify and harmonize data in the context of a specific use case.

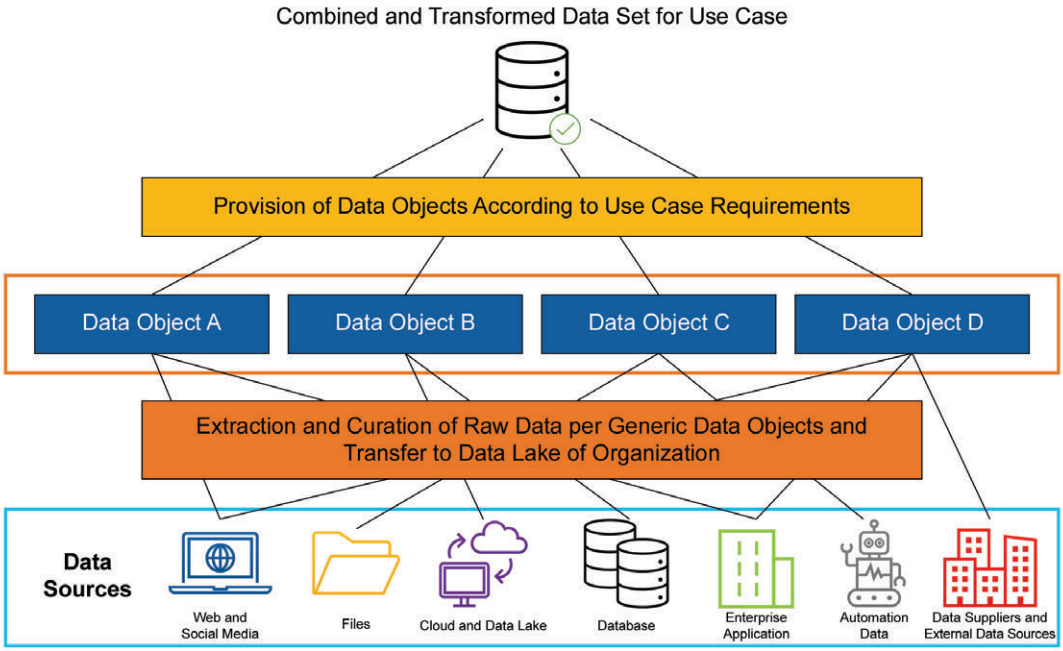
Figure 18.3: Illustration of Data Collection Strategy with Global Data Storage



Data Collection Strategy with Global Generic Data Objects

This strategy involves designing generic data objects and mapping data source contents to these data objects during integration; see Figure 18.4. Curation and transformation are already carried out before the data is stored in the data storage. While regulated companies bear initial effort to establish generic data objects, they can take advantage of simplification and cost savings when the data is used for a variety of use cases.

Figure 18.4: Data Collection Strategy with Global Generic Data Objects



18.4 Data Access and Provision

Data provisioning techniques involve batches or automatized pipelines or streams. Suitability depends on the purpose of data provisioning. For instance:

- Model engineering in the project phase is based on historical data which is typically provided in batch form to construct the case data set; see Appendix P2.
- In the operation phase, data streams or pipelines are commonly used to allow for ongoing monitoring and further use of data; see Appendix P3.

Data Batches

Data batches are collections of data records according to specific filter criteria, thereby available as a complete, fixed set. This is useful for data analysis and for prototypes, see Appendix P1, or during model development, see Appendix P2.

Records on the choice of filter criteria allow for traceability of technical artifacts, such as models or performance indicators, and reproducibility (depending on the model type). In addition, use of the exact same data allows for targeted iterative experimentation when developing models, excluding the influence of changing data sets, unless more data is purposefully selected during iterative experimentation.

Data Pipelines or Streams

Data pipelines or streams directly connect data sources via connectors. Data is read from the source as required or sent by the source for specific events. Depending on the source system type, the connectors can be designed to handle different types of data transfer, such as:

- Database connectors

- Message connectors
- Communication connectors (preferably compliant with communication standards)

Connectors are often linked to parameterization options that can be used to define rules for selecting tables, data records, and attributes during data selection.

Automation of data pipelines or streams allows for effective use of data, for instance, for ongoing monitoring purposes. Integration with tools such as real-time dashboards can provide summarized information to derive actions.

18.4.1 Data Transfer and Communication Between Systems

Synchronous and asynchronous data transfer techniques can be used for data transfer, influencing the timeliness of when data becomes available in a respective other system:

- **Synchronous transfer:** Every change in the source is immediately transferred to the target system. Data is therefore available in real time or near real time. The disadvantage is that it can place a high load on the infrastructure and systems.
- **Asynchronous transfer:** This method involves extraction independent of changes in the source systems. It can take place periodically, be event-driven, or on request.
  - **Periodically:** A source system provides data at regular intervals, which is regularly queried, or an extraction process reads in data at regular intervals.
  - **Event-driven:** A source system provides data when certain events occur (e.g., after a certain number of changes), or an extraction process reads in data on an event-driven basis.
  - **Request-driven:** A source system only provides data on request, or an extraction process reads in data on request.

Data transfer modes should be selected based on the following factors:

- System capabilities (e.g., the interfaces provided by software products)
- Infrastructure considerations
- The needs of data in the target system to fulfill its intended use
- Risks of possible latency of data transfer

18.5 Data Cleansing and Data Transformation

After data extraction from a source, the data can optionally be prepared for further use. In addition to improving data quality, the aim of the transformation is often to simplify the handling of the data in subsequent steps.

Several steps are commonly used when performing data cleansing and transformation:

- Dealing with missing data
- Handling outliers
- Correction of errors in data

- Labeling data with its semantic meaning
- Transformation into a form suitable for AI

18.5.1 Syntactic Transformation of Data

Various syntactic transformation methods help ensure that data adheres to formal aspects and is modified if necessary. Examples of the transformation steps include:

- **Adaptation of data types:** For example, values may be in text form although representing numerical values, and may be converted into a suitable data format.
- **Display of special formats:** Some types of data may vary in their representation; for example, representation of date and time frequently varies due to the place of origin. Standardization, especially time zone adaptation, is essential for such cases.
- **Interpretation of codes:** Examples include different status coding (e.g., 1/2/3 or l/m/h for low/middle/high), different country codes, or the use of comma versus period to denote decimal values.
- **Adaptation of units of measurement:** Data often only acquires a semantic meaning through units of measurement, e.g., the number 5 is only a natural number, but the unit kg indicates it is a weight. If the data that is to be analyzed in context is available in different units, it should be adjusted through conversion (e.g., 5 kg becomes 5000 g).
- **Alignment of measurement precision:** Values may have been recorded with different precision over time or measured by different equipment (e.g., scales with different accuracies). For example, the accuracy of a weight measured by two different scales may vary.

18.5.2 Handling of Duplicates

Duplicates may occur when collecting data from various sources, i.e., data records with identical or close content, representing the same real-world entities. Since duplicates may have negative consequences on models, such as by overrepresenting certain cases, regulated companies should develop strategies to identify and mitigate duplicates.

Examples for handling duplicates include metadata analysis, fuzzy matching (i.e., techniques to match records that are close, though not exactly, matching), and advanced techniques:

- **Using metadata for duplicate identification:** Metadata, such as batch numbers, timestamps, and facility IDs, can help identify duplicate records. For example, if two records share the same batch number but differ in minor details, they may represent the same event, though with differences in accuracy of information. Analyzing metadata such as timestamps and production conditions helps verify potential duplicates.
- **Fuzzy matching techniques:** Fuzzy matching is useful when entries are similar but not identical. Common techniques include, for example:
  - **Levenshtein Distance:** Measures text distances, which can be used to identify misspelled names (e.g., “Jon Doe” vs. “John Doe”).
  - **Phonetic Matching:** Flags similar-sounding names (e.g., “Atorvastatine” and “Atorvastatin”).
  - **Jaccard Similarity:** Compares the similarity between two sets of values. For instance, it can be used to compare ingredient lists or formulations with minor formatting variations.

- **Advanced techniques:** ML techniques can help identify duplicates by learning patterns from historical data, including use of human verification for training purposes to derive similarity measures, and suggested duplicates based on these measures.

18.5.3 Handling Missing Values

Various techniques can be used to address missing data and thus provide more robust foundations to achieve effective models. Techniques range from simple replacement to advanced model-based approaches, to handle missing values:

- **Simple replacement techniques:**
  - **Mean, median, or mode replacement:** Replace missing values with the mean, median, or mode of the available data. This approach works well for small gaps but can introduce bias if data is missing systematically.
  - **Interpolation:** Fill in time-series data using interpolation (e.g., linear or polynomial) to estimate missing values, based on trends in adjacent data.
  - **Labeling missing values:** Explicitly mark missing values, helping the model recognize gaps and distinguish them from errors.
- **Advanced replacement techniques:**
  - **Nearest neighbors:** Use the nearest neighbors of incomplete records to estimate missing values, based on a suitable metric.
  - **Multiple replacement by chained equations:** Generate multiple replacements for missing data based on relationships between variables, preserving data variance.
  - **Predictive models:** Use ML models to estimate or predict missing values based on patterns in the data set.

18.6 Data Augmentation

Data augmentation techniques are employed to generate additional data, see Appendix P2. They help to extend coverage of the expected data space, thus improving representativeness of data expected in operation. Another use case includes assessment of limitations of model performance via scenarios that are rarely observed. Data augmentation techniques include:

- **Geometric transformation:** Algorithmic transformations (e.g., rotations of images) can help expand the data set.
- **Physics-based models:** In some cases, the applicable data space can be described by physical models (e.g., from engineering or biology) which may be used to create synthetic data.
- **Synthetic data generation:** Models such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs) can be used to create synthetic data.
- **Scenario-based data creation:** Depending on the complexity of data, it may be feasible to construct specific, plausible scenarios. They may be used to cover rare events otherwise not present in data.

In addition, **bootstrapping** (resampling the data to generate new, varied samples) can be a useful technique when working with missing or sparse data.

Any model used for replacement or augmentation should be assessed regarding its fitness for purpose. This may include assessing how replacement or augmentation techniques impact model accuracy by comparing different methods on the same data set.

18.7 Model Management

The technical representation of a model ("model artifact") should be captured in a model inventory, similar to maintaining data inventories for data. This facilitates traceability of origin and use of models, and contextualizing models with supportive metadata.

Descriptions and references to model development activities (see Appendices P1 and P2) support assessment of suitability of models for a specific case. For instance, such descriptions and references may help to decide between direct use for refining a base model, in conjunction with an understanding of the use case.

Maintaining linkage between data and models should be ensured. This includes its use for various purposes such as model development and testing of AI sub-systems to ensure transparency of decisions. This helps to substantiate the suitability of a particular model regarding the representativeness of the underlying data for the target use case. Similarly, capture of model performance indicators derived during the evaluation and testing of models complements information on the model development process.

Code that has been used to engineer models should be captured in a way to allow full traceability and reproducibility of the model, depending on model characteristics. Good software engineering practices and DevOps principles as well as suitable IT infrastructure are helpful to achieve this goal.

Changes or deployments should be captured via change management processes in the model inventory. Similarly, decisions regarding the retirement of AI-enabled computerized systems or its AI sub-systems should be considered to maintain an up-to-date state of usage of models.

Further technical aspects are provided in Appendix M8.

18.8 Data Operations and Implications on Models

Four basic data operations can have implications on models:

- **Create:** When data is created the first time
- **Update:** When data is updated, e.g., when data needs to be corrected or the value changes
- **Select:** When data is retrieved
- **Delete:** When data is removed from storage

In the case of GxP data, data operations are governed by a strict life cycle to ensure that data integrity expectations are met. See *ISPE GAMP Guide: Records and Data Integrity* Chapter 4 [15], which outlines the following key phases:

- Data Creation ("Create")
- Data Processing ("Select")
- Data Review, Reporting, and Use ("Select" and "Update")

- Data Retention and Retrieval (“Select”)
- Data Destruction (“Delete”)

For non-GxP data, less strict processes may suffice, while similar activities leveraging data operations are typically executed on that data.

Regardless of whether GxP or non-GxP data is handled, data ownership should be defined. This includes outlining who is responsible for data at different stages and ensuring that departments or teams understand their roles and responsibilities.

Data operations may have implications on the respective model. Activities such as model development may reveal potential shortcomings in the way data is managed and may consequently lead to improvements in the data management process and corrections in case of errors.

A systematic approach that links data management and model management activities help mitigate common risks such as:

- Poor data quality
- Model drift
- Bias of data

These can all lead to inaccurate predictions or decisions that introduce risk to patient safety, product quality, and data integrity.

In the following sub-sections, considerations on key data operations and their implications for models, AI sub-systems, and AI-enabled computerized systems are described in more detail.

18.8.1 Data Creation

Data creation is the process of generating new data, either through direct collection, simulation, or transformation of existing data. Regulated companies may create this data for maintaining records of a process, or for use in analytics, ML, or other computational systems.

How data is created and how its quality is ensured is highly relevant to ensuring its suitability for an AI use case. Regulated companies should consider the following aspects when creating data and when evaluating whether a data set has been created under acceptable circumstances:

- **Human quality assurance of data:** In situations where domain specificity is high or context is required, human reviewers can verify the correctness of the generated data, depending on the complexity of data. Regulated companies should establish common practices to harmonize human quality assurance of data.
- **Automated quality assurance:** Automated quality assurance can be applied to data by implementing rule sets, such as checking for expected formats or value ranges, highlighting the presence of null values, or checking for inconsistent or implausible data constellations in respective interfaces before they are recorded. Warnings can be raised when such rules are triggered.

  Employing automated quality assurance techniques and rules that are relevant for the process should be based on product and process understanding, and data understanding.
- **Monitoring of generated data:** Processes to regularly reviewed generated data can help detect changes in patterns or distributions, thus spotting potential deficiencies in the way data is created early on; see Appendix P3.

- **Review of data creation practices:** Regular reviews help to ensure that data creation processes meet industry standards and regulatory requirements, as applicable; see Appendix P3.

These considerations help maintain the reliability of the data and qualify it for potential selection in the development of models, considering the suitability for their context of use.

Conversely, the requirements of AI-enabled computerized systems may necessitate changes to the way data is created. For example, regulated companies may consider adding sensors to capture more data in a manufacturing setting, thus enriching a model with further real-world characteristics of the process.

18.8.2 Data Updates

Changes to data may occur as errors are detected or as more is learned about what the data should represent. Data quality assurance and anomaly detection help identify and correct errors before data is used for analysis or model development, evaluation, or testing, thus leading to data updates.

Changes should allow for traceability of the data that was used for model engineering, evaluation, or testing. Options range from simpler data version management techniques to more formal approaches, such as establishing a formal audit trail for GxP data.

Data management should be considered at the level of organization, as multiple uses of data for various models may be possible. In this situation, data corrections may have substantial implications for the validation state of all AI-enabled computerized systems that embed AI sub-systems, and models, that are linked to erroneous data. Therefore, changes should be reviewed to decide on performing an impact assessment regarding the relevance of such changes for models linked to that data.

On the other hand, model development activities may expose shortcomings in the data, which in turn can trigger necessary updates to the data according to established rules and policies. Further steps and actions may be necessary to prevent shortcomings in the future, particularly regarding GxP data.

18.8.3 Data Selection

Data selection in the context of AI-enabled computerized systems typically occurs when establishing a prototype for a PoC (see Appendix P1) or constructing the case data set (see Appendix P2). Here, teams typically collect data from various sources; see Section 18.3.2.

The following aspects are important when selecting data:

- **Data access** should be restricted to authorized personnel, with Role-Based Access Control (RBAC) in place to ensure that sensitive data is only accessible to those with appropriate clearance.

- **Data should be assessed for its fitness for purpose.** Benchmarking and comparison to external data can support the identification of inconsistencies or discrepancies early. Further guidance on EDA is in Appendix P1.

- **Reproducibility** of the extracted data set should be ensured. For instance, versions to data sets may be attached, with a protection of changes, or reproducibility can be achieved by capturing metadata of data collection, such as the point in time of extraction and applied filter criteria.

Data selection is in principle neutral to a computer system, while infrastructure load and cybersecurity aspects should be considered. However, the analysis of such data may lead to corrections or updates to the original data set. In such cases, regulated companies should reprocess data for the purposes of model development activities (recommended option) or correct such data in a way that is equivalent to the corrections introduced in the source system (applicable option, though it carries the risk of deviations between both data correction approaches and may impose a burden on maintenance of data integration code and future model development activities).

18.8.4 Data Deletion

Data deletion involves the removal of data, i.e., the permanent and irreversible erasing of that data from all systems. In a GxP setting, data deletion is governed by defined retention policies that outline how long data should be stored in compliance with applicable regulations (see further guidance in *ISPE GAMP RDI Good Practice Guide: Data Integrity by Design* Table 16.1 [1]) and statutes such as personal data protection laws to protect privacy.

Deletion of non-GxP data should be performed with caution, considering further aspects such as securing IP, if relevant for the context. Based on an evaluation, a decision is made whether to archive some data for use in *ex post* analysis, research purposes, or future use for model development. (See Appendix P4.)

Governance policies support distinguishing between archiving (where data is stored securely but remains accessible) and deletion (where data is permanently destroyed). The full traceability of a model may be lost if data used for development, evaluation, or testing purposes is deleted. Alternative measures, such as anonymization or pseudonymization, can be applied to keep essential information used in conjunction with the model. The possibility of re-identification (i.e., identifying the individual from non-altered data) needs to be considered and, as applicable, mitigated.

18.9 Roles and Responsibilities

An organizational layout of roles and responsibilities supports data and model governance and management activities. In the context of AI-enabled computerized systems, various levels apply:

- **Organizational level:** Organizations should consider data and models from an organizational perspective. They should be aware of potential further use of data and models to support other use cases, demonstrate generalizability of AI approaches, or leverage AI approaches across different contexts of use.

  Data stewards support integration of knowledge about data at the organizational level, helping interested parties navigate the data landscape and ensure adherence to the organization's data quality expectations. Similarly, model stewardship may be defined and established accordingly. Thus, stewards also support knowledge management activities, as more insights on the use of AI are collected across the organization.

  A general counsel or governance board may support such stewardship, determining the appropriateness of the organization's practices and deciding on initiatives to strengthen data and model governance and management practices.

  Furthermore, alignment of incentives regarding data and model governance and management can be underpinned by a corporate data and model strategy that implements incentives to achieve long-term benefits from high-quality, well-understood data pools and model repositories.

- **Process and system level:** Data and models fulfill a purpose within the AI-enabled computerized system, where data is used for development, testing, and monitoring of the model's performance. Data owners, in some cases coinciding with the process owner, and system owners need to be aware of the data type generated, and how models are integrated into the computerized system and the GxP regulated process.

- **Sub-system level:** Owner roles need to manage the development, use, and further development of the sub-system, either as a designated role or as part of the data owner's responsibilities. Sub-system owners should take advantage of data generated as part of the process and further data collected within the organization, as applicable.

To foster these activities in regulated companies, the Quality Unit takes an oversight position, supporting adherence to organization's quality expectations regarding data and model governance and management. In addition, the Quality Unit also supports the integration of supplier activities, addressing data and model governance and management considerations; see Chapter 6 and Appendix M2.

18.10 Scaling of AI Use Cases

Scaling AI use cases within the organization means fitting additional computerized systems with established use of AI elsewhere in the organization or even collaborating with organizations that choose to partner in leveraging AI capabilities. Points to consider include:

- Learnings from implementation in one area can foster further utilization. For example, using insights from testing of AI-enabled computerized systems in one area can improve the integration of the model or upskilling and training of staff in other areas.

- Data re-utilization and model inheritance may offer options such as utilizing pooled data or fine-tuning of existing models. This may be favorable from a perspective of efforts, underpinned by a business rationale and an assessment of representativeness of data used and generated by AI-enabled computerized systems in operation in other environments.

  In addition, similar to the use of foundation models provided by suppliers, further complexity is expected in life cycle management if models are interlinked, which should be justified by appropriate benefits.

A strategic approach to the dissemination of a use case throughout the organization supports preparing for scaling early on, if applicable. For example, if a use case has a high possibility of dissemination across the organization, more effort may be taken into automating processes.

18.11 Data Sharing and Federated Learning

Data sharing and federated learning provide promising approaches to improve or to test the generalizability and robustness of models. Federated learning approaches aim to share model artifacts, to arrive at more generalized combined models without sharing the original data. In contrast, data sharing involves the actual transfer of data, e.g., from a data provider or a regulated partnering organization, for agreed upon activities.

The legal and contractual implications of such activities should be considered, in addition to the following points:

- Fitness for purpose of data needs to be assessed, with specific consideration of representativeness of data for the use case at hand. A rationale for the chosen approach should outline the expected benefits while assessing the specific risks.

- Benefits of data sharing or federated learning should be measured, allowing for risk-based decision-making. While an assessment of data is possible in the case of data sharing, the use of federated models requires more trust in the individual parties' contributions, adhering to an aligned set of characteristics and agreeing on an adequate level of transparency into data distributions and statistics. In both cases, openness and transparency regarding possible shortcomings in data or models is required.

- Similar to the involvement of foundation models when cooperating with suppliers, an increased complexity of model management activities is typical when shared data or federated models are involved.

- Risks may arise from erroneous data introduced by other organizations. There should be increased awareness of their accountability in impacting other organizations.

Organizations may agree to share results from risk and quality management activities to allow for quick remediation of potential weaknesses when considering federated learning approaches.

19 Appendix M8 – IT Infrastructure

19.1 Introduction

A controlled IT infrastructure is a prerequisite for ensuring that GxP applications are managed in a state of control. Infrastructure plays a significant role in ensuring computerized systems are on a stable platform. [2] IT infrastructure supports performance and availability as well as the integrity, availability, security, and confidentiality of data. Many of the processes required to ensure this rely on some aspects of infrastructure management, such as cybersecurity, load balancing, backup and restore, disaster recovery, etc. [2]

In addition to considerations provided by *ISPE GAMP 5 (Second Edition)* [2], this appendix provides an overview of typical infrastructure elements that support the management of data and models. Models, like other technical artifacts like data or configuration and software code, require careful handling, inventory and version management to allow for traceable use and possible reuse. In addition, regulated companies should consider relations between technical artifacts like models and data.

In managing data, software code, and models for AI-enabled computerized systems, several types of technology such as software products or tools are available to streamline processes, support robust governance and other data and model management activities.

Further guidance on related concepts like the use of tools, DevOps, Continuous Integration and Continuous Deployment can be found in *ISPE GAMP Good Practice Guide: Enabling Innovation* [85], while guidance on IT Infrastructure Control and Compliance can be found in *ISPE GAMP® Good Practice Guide: IT Infrastructure Control and Compliance (Second Edition)* [111].

This appendix addresses regulated companies, following the assumption that they maintain MLOps architectures for their own development and operation purposes. However, suppliers of software products, utilizing similar MLOps architectures for their operations, may utilize guidance provided here to understand typical expectations of regulated companies on their practices.

Regulated companies should take a holistic perspective on the management of data, software code, and models in the context of AI-enabled computerized systems. Various tools can be integrated to form an MLOps (ML Operations) architecture, allowing:

- Teams to effectively manage data, software code and models throughout the system's life cycle
- Smooth transitions from experimentation to production
- Continuous monitoring and improvement
- Reuse of data, software code and models

Options in establishing an MLOps architecture include combining a set of tools, applications, and services, and utilization of configurable MLOps platforms that provide a combination of services.

Generally, regulated companies should consider aspects such as capture of records and information, traceability, maintainability, and good software engineering practices (e.g., code and configuration review) for technical artifacts. These practices should be commensurate with the risk. Third-party software should be actively managed as described in *ISPE GAMP 5 (Second Edition)* [2], particularly Appendices D9 and M11. Infrastructure software is classified as GAMP Category 1, and as such expected to be qualified, not validated [2], while further information on infrastructure qualification can be found in *ISPE GAMP 5 (Second Edition)* Appendix M11 Section 19.5, covering:

- Security Management

- Backup and Recovery
- Archive and Restore
- Change Management
- Configuration Management
- Disaster Recovery and Business Continuity Planning
- Other IT processes

Cloud infrastructure, including variants such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) are also relevant for MLOps architectures managing data, software code and models in the context of AI-enabled computerized systems. Reasons include the availability of high computing power, when necessary, large data storage or specialized hardware, such as GPUs.

Further considerations on risk management activities and the role of Quality as well as continual improvement apply as per *ISPE GAMP 5 (Second Edition)* [2].

Common practices across projects and operational use of AI-enabled computerized systems should be established, unless a rationale for a specialized is present (e.g., exclusive provision of certain models via a dedicated software product or platform).

The following subsections provide an overview of typical MLOps architecture elements.

19.2 Data Storage

Data storage refers to infrastructure elements dedicated to storing and organizing data. Accessibility, reliability, and integrity of data should be ensured over time.

A robust data storage architecture is essential to achieve effective and safe AI-enabled computerized systems, considering potentially large volumes of sensitive data and the need to protect patient information.

Typical choices of data storage layouts include:

- **On-premise storage:** Data is stored on local servers, often within the organization's physical infrastructure. This option offers control over data and security but requires significant maintenance and hardware costs and poses scalability challenges.
- **Cloud storage:** Data is stored in the cloud using external service providers. Cloud storage offers scalability, flexibility, and reduced infrastructure costs while providing advanced security and compliance support. However, cloud storage may lead to latency, which may be inadequate for close to real-time applications like process control systems.
- **Hybrid storage:** This type combines on-premise and cloud storage. Critical or sensitive data may be stored on-premise for performance, security, or compliance reasons, while less sensitive data or data requiring frequent access is stored in the cloud. This model offers a balance between control and flexibility.

A suitable layout for data storage can rely on various options:

- **Data lakes:** Used for storing large volumes of structured, semi-structured, or unstructured data. Data lakes store raw data that can be processed later for various purposes (e.g., analysis, reporting, ML).

- **Data warehouses:** Structured and processed data is stored in data warehouses for easier querying and reporting. They are commonly used to aggregate and analyze data.

- **Distributed network architectures:** Data storages organized as distributed network architectures, including variants like blockchain implementations, allows for distributed data management, with additional features such as verified credentials and embedded cryptographic techniques. Such architecture may be helpful to manage widely spread data, with the capability to demonstrate data integrity as well as for traceability of data in cross-organizational settings.

Relevant requirements for data storage include:

- **Scalability and performance:** Large volumes of data may be generated throughout the lifetime of an AI-enabled computerized system. Data storage solutions should be scalable to accommodate growing data sets while maintaining high performance for data retrieval, processing, and analysis.

- **Security:** Data in the context of AI-enabled computerized systems may be highly sensitive. Therefore, data storage solutions should include robust security measures, including:

  - Encryption: Both at rest and in transit, encryption ensures that data is protected from unauthorized access.

  - Access Control: RBAC ensures that only authorized personnel have access to specific data sets, minimizing the risk of data breaches or accidental data exposure.

  - Data Backup and Recovery: Regular backups and availability of effective recovery mechanisms are essential for safeguarding data against loss due to hardware failure, cyberattacks, or other incidents.

- **Compliance:** Data storage should comply with various regulations that define how data needs to be stored, accessed, and preserved over time. Key regulatory frameworks include:

  - 21 CFR Part 11 [89]: Governs the storage of electronic records and requires that electronic systems used in drug development ensure data integrity, traceability, and security.

  - GxP: These guidelines outline specific storage requirements and guidance for clinical (e.g., Guideline on computerised systems and electronic data in clinical trials [112]) and manufacturing data (e.g., EU GMP Annex 11 [113]), ensuring that data is preserved for the required retention periods and remains accurate and accessible for audits.

The fulfillment of these requirements should be verified as part of determining data being fit for purpose.

The following aspects are of further consideration when managing data storage to support one or various AI-enabled computerized systems:

- **Management of records and information** as evidence that the storage is fit for purpose.

- **Classifying and segmenting data** based on its sensitivity, regulatory requirements, and business use. For example, patient data, regulatory data, and IP should be classified into different categories with varying levels of protection and retention requirements.

- **Organizing strict access control,** ensuring that only authorized personnel can access sensitive data. These policies should be regularly reviewed and updated.
- **Defining retention and deletion schedules** for each category of data.
- **Documenting audit trails** to track who accessed, modified, or transferred data.
- **Upgrading technologies or storage solutions** with data migration plans that ensure data remains intact, secure, and compliant with regulatory requirements.

19.3 Data Catalog and Metadata Management

Leveraging a centralized repository to catalog and manage metadata related to data assets ensures seamless access and allows discovery of data for development teams. It enables them to explore data sets and understand their origins, relationships, and governance policies to decide on the suitability of such data for their use case.

Common features include:

- Automated metadata captured from data sources, applications, and pipelines
- Data discovery tools that allow users to search and explore data sets across the organization
- Data lineage visualization to track the flow of data through different stages in the pipeline
- Data governance features, such as tagging, ownership assignment, and access control

19.4 Data Collection and Ingestion

These tools and software products are designed to ingest data from a variety of sources (such as databases, APIs, sensors, or logs). They support real-time or batch data ingestion and handling of structured and unstructured data, including automated scheduling of data extraction, transformation, and loading (ETL) processes.

Common features include:

- Data connectors for various sources (e.g., relational databases, APIs, IoT devices)
- Support for various file formats (CSV, JSON, XML, etc.)
- Scalability for handling large data sets
- Real-time streaming capabilities and event-based triggers

19.5 Data Pre-Processing and Transformation

Tasks supporting data pre-processing and transformation involve cleaning, normalizing, and transforming raw data into a format suitable for analysis or model engineering. This includes automated data quality assurance and error detection, data normalization, aggregation, and feature engineering and handling of missing data, outlier detection, and type conversions.

Common features include:

- Data profiling to assess the quality and completeness of the data
- Pre-built and customizable functions for data cleaning, such as removing duplicates and handling missing values
- Support for joining, merging, or aggregating data from multiple sources
- Integration with ML workflows for automatic feature engineering

19.6 Data Versioning

Data versioning tools and software products allow for management and capture of the history of data sets, thus ensuring reproducibility and traceability. Data lineage capabilities allow teams to track the origin, changes, and destination of data sets.

Common features include:

- Version control for data sets, such as how the software code is version controlled
- Metadata management and data cataloging to provide descriptions, relationships, and governance rules for data sets
- Access control and data usage auditing for compliance with privacy and security regulations
- Data lineage tracking to visualize and audit changes to data sets over time

19.7 Data Security and Compliance

Data security and compliance solutions focus on ensuring that data is stored, managed, and processed securely, adhering to industry standards and applicable regulations. This includes supporting encryption, access control, and audit logging to protect sensitive data.

Common features include:

- RBAC and user authentication to manage data access
- Data encryption both at rest and in transit to prevent unauthorized access
- Audit trail and reporting to track data usage, ensuring accountability and transparency
- Compliance with industry regulations such as 21 CFR Part 11 [89] or EudraLex Good Manufacturing Practice Annex 11: Computerised Systems [113]

19.8 Data Pipeline Orchestration

The purpose of data pipeline orchestration is to automate and orchestrate end-to-end data workflows, such as data ingestion and pre-processing or data engineering, or when used for monitoring purposes. This supports the creation and scheduling of complex workflows (e.g., multiple steps with dependencies among steps), ensuring that processes are executed in the correct sequence while allowing for parallel processing, where applicable.

Common features include:

- Workflow scheduling and automation, supporting both time-based and event-based triggers
- Support for integrating data and ML pipelines, ensuring that data flows seamlessly from preprocessing to data modeling
- Error handling and retry mechanisms to ensure reliability and fault tolerance in workflows
- Visualization of data pipelines for monitoring progress and identifying bottlenecks

19.9 Model and Artifact Management

Centralized model and artifact storage can support the management of technical artifacts of AI sub-systems, such as models, data sets, and code. This ensures the versioning, sharing, and security of such artifacts across teams.

Common features include:

- Artifact repositories for storing models, source code, parameters, and data sets with adequate version control
- Model packaging and containerization to ensure reproducibility and portability across environments
- Security features to protect artifacts, ensuring access control and encryption
- Collaboration tools to share and reuse models or data sets across teams

19.10 Tracking of Iterative Experimentation

Software products and tools for tracking iterative experimentation help to capture configurations, data, and model performance (see Appendix P2), and can facilitate the comparison of the performance of various models.

Common features include:

- Training cycle tracking to capture model parameters, metrics, and artifacts from multiple training runs
- Versioning of models, code, and data sets to ensure reproducibility
- Automated model engineering pipelines, enabling hyperparameter optimization and evaluation
- Integration with ML libraries and frameworks

19.11 Model Monitoring and Drift Detection

Software products and tools for model monitoring and drift detection help to verify that models perform as expected in operation. This allows for model drift to be detected; see Appendix P3. Alerts or automatic retraining triggers may be included if performance thresholds are breached.

Common features include:

- Real-time tracking of result accuracy, latency, and resource consumption
- Model drift detection through statistical comparisons between training data distributions and incoming live data
- Automated alerts and notifications for model performance degradation or failures
- Integration with retraining workflows to ensure models adapt to changing data environments

19.12 Model Deployment and Serving

Exposing models via Application Programming Interfaces (APIs) can facilitate the deployment of models into production environments. While models may also be directly embedded in the software of the AI-enabled computerized system, the modularized deployment of models can facilitate the iterative transition of models to new versions.

Common features include:

- APIs or microservices to allow for integration of models for inference in real time or for batch processing purposes
- Support for model versioning and rollback to deploy new models and manage changes
- Resource management to scale models based on usage, ensuring efficient performance under varying loads
- Integration with Continuous Integration and Delivery (CI/CD) pipelines to automate deployment workflows

20 Appendix M9 – Trustworthy AI

20.1 Introduction

This appendix expands on the trustworthy AI principles introduced in Chapter 2. Trustworthy AI considerations are recognized by health agencies. For example, the EMA reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle [43] states that for drug discovery, *"all models and datasets used should be reviewed by the sponsor to mitigate ethical issues, risks of bias and other sources of discrimination of non-majority genotypes and phenotypes from a data quality and quantity perspective."*

This appendix contextualizes these principles in the life sciences, linking to further concepts provided in this Guide and further resources. It seeks to support regulated companies in the implementation of Trustworthy AI principles in their life cycle activities and help suppliers anticipate the expectations of regulated companies.

Decision-making regarding trustworthy AI should rely on a collaborative effort that includes diverse stakeholders and should be underpinned by a rationale and ideally supported by evidence.

Considerations on trustworthy AI are more pronounced in some areas than in others; for example, fairness may be a major concern for the use of AI in medical devices or clinical trials, while such considerations may be less prominent for commercial manufacturing of small molecule pharmaceuticals.

20.2 Human Autonomy and Control

- Choice of an adequate level of control and allowing for the possibility of humans to regain control over the AI-enabled computerized system as needed; AI maturity offers various choices on the level of autonomy of an AI sub-system (see Appendix M10)
- Contextualization of typical end-user capabilities and planning of training to allow for adequate human oversight (see Appendix M5)
- Impact of AI on human decision-making: AI should enhance and complement, rather than replace, cognitive decision-making and should not induce pressure that may yield an incorrect decision [28]
- Relationship between oversight and control, and impact on patients, see WHO [114]
- Using XAI methods to support users in their control function, as applicable by choice of the system's autonomy; see the FDA/Health Canada/MHRA guiding principles [115] and Appendix S4
- Fostering trust by establishing human control points throughout the process and, if applicable, during operation; see Appendices P2 and P3
- Monitoring of changes in human oversight during use with unfavorable implications on the process, e.g., rigor and vigilance may drop, or control may be executed excessively albeit not justified; see Appendix P3

20.3 Safety and Security

- Establishing AI literacy and awareness for security risks across the organizations and for end users; see Appendix M5
- Promoting risk management activities concerned with impact, and thus safety of patients to derive suitable controls; see Appendix M3
- Providing a rationale that underpins the evaluation of the benefit-risk balance of choice of AI; see Appendix M3
- Making a thoughtful choice of the system's maturity level, based on the context of use, the organization's experience, available data, and risks; see Appendix M10
- Designing and selecting models considering the sensitivity of the model to its input; see Appendix P2
- Using testing strategies that cover the performance, robustness, and stability, as well as cybersecurity resilience of AI-enabled computerized systems; see Appendix P2
- Considering cybersecurity aspects, including AI-specific threats and tracking of potential attacks; see Appendix S5
- Planning business continuity processes, including notification of relevant stakeholders; see Appendix P3

20.4 Fairness and Mitigation of Bias

- Interpretation of fairness in the context of use; for instance, use of AI in clinical trials or in medical devices may provide more benefits to some sub populations than others; see Appendix P1
- Reflecting on experience, cultural, social, historical, political, legal, and further ethical factors; see Appendix P1
- Identifying relevant dimensions that provide stratifications of patients for which sufficient evidence on the performance of models and systems needs to be established; see Appendix P2
- Assessing possible implicit biases in data or models, as identifiable via analysis of data or when evaluating data, labeling and model development practices; see Appendix P2
- Considering possible amplification of bias, when bias from data is perpetuated in models, which produce further model output; see Appendix M7
- Monitoring of potential bias during operation of the AI-enabled computerized system, with established human oversight to evaluate impact of potential bias; see Appendix P3
- Use of XAI methods that may help identify underlying patterns that exhibit bias; see Appendix S4

Oversampling and Use of Metrics That Are Insensitive to Class Imbalances

As suggested by EMA in their reflection paper [43], *"the need to over-sample rare populations should be considered, taking all relevant bases of discrimination as specified in the EU principle of non-discrimination and the EU fundamental rights into account."* Various techniques can be used to shift the weight of observations during model engineering, including sampling from the case data set with different proportions, for instance by sub-populations, or applying higher weights in the optimization algorithm to observations that are underrepresented. The choice of the parameters, and the implications on the model should be carefully assessed to verify that indeed these techniques supported reduction of bias.

Further to that, EMA [43] also mentions the potential choice of metrics that are insensitive to class imbalances, which constitutes a way to evaluate models considering potential underrepresentation of data such as particular populations or rare events.

20.5 Privacy and Data Protection

- Raising awareness for privacy and data protection within the organization
- Considering applicable statutes and regulations on a local and regional level, such as HIPAA [26] or EU GDPR [25]; see Appendix P1
- Ensuring adequacy of consent when personal data is used for AI-enabled computerized system development; see Appendix P1
- Considering the implications of processing sensitive data on individuals, e.g. health data generated in clinical trials or when used for clinical decision-making; see Appendix P1
- Establishing robust system and data infrastructures; see Appendices M7, M8, and S5

20.6 Transparency

- Ensuring awareness for those interacting with AI; see Chapter 6
- Ensuring sufficient AI literacy and provision of sufficient training on system-specific aspects of the AI-enabled computerized system; see Appendix M5
- Promoting science-based decision-making, grounded in evidence and risk considerations; see Appendix M3
- Transparency regarding information on the benefits, risks, and limitations of the use of AI; see the FDA/Health Canada/MHRA guiding principles [33]
- Provision of sufficient information on development practices of AI-enabled computerized systems, their architecture, and use of data; see WHO guidance "Regulatory considerations on artificial intelligence for health" [114]
- Evaluating the suitability of XAI methods that support interpretability of model output; see [116]
- Provision of adequate level of detail when conveying information; in particular *"who (relevant audiences), why (motivation), what (relevant information), where (placement of information), when (timing) and how (methods used to support transparency)"* [33]

- Adoption of FAIR principles (Findable, Accessible, Interoperable, Reusable), involving well documented, easily accessible, and interoperable data and models; see Appendix M7
- Ensuring traceability between model input, models, model output, and decisions; see Appendix M7

Transparency by Design

As outlined in the FDA's draft guidance [102], transparency by design fosters a holistic approach regarding transparency, *"throughout the full continuum of implementation through use, maintenance, and decommission of the AI-enabled device,"* including "transparency in mind from the beginning." Not restricted to devices, capturing relevant information and maintaining an overview of various aspects and challenges throughout the life cycle is important for presenting information and records comprehensively to achieve adequate transparency for stakeholders.

Transparency of Use of AI in Clinical Trials

When deploying AI for use in clinical trials, consideration needs to be given to inform the participant, e.g., via informed consent forms or information on clinical trial protocols. If an AI-enabled computerized system is used to make or support decisions affecting the participants, then this should be transparent to them to enable an informed choice about enrolling in the trial. If an AI-enabled computerized system is utilized in patient-facing applications and clinical trial protocols (such as an electronic patient reported outcome (ePRO) or chatbots), then it should be clear to the participant that they are interacting with AI (refer to Section 20.4). For further information on the use of AI in the clinical trial context, and specific considerations on clinical trial stakeholders, refer to the *ISPE GAMP® Good Practice Guide: Validation and Compliance of Computerized GCP Systems and Data – Good eClinical Practice (Second Edition)* [58].

20.7 Accountability

- Establishing clear roles and responsibilities across the organization and in case of supplier involvement; see Chapter 6
- Establishing appropriate human oversight to fulfill accountability; human oversight does not necessarily mean verification of each model output, see also Appendix M10
- Raising awareness of the decisions of individuals; for instance, end users need to be aware of their oversight role, while technical roles should consider implications of design choices of complex models; see Appendices P3 and M5
- Raising understanding of potential impact of the use of AI in the context of use; see Appendix M3
- Employing scalable life cycle activities, based on critical thinking; see Appendix M4
- Monitoring of and adherence to regulatory requirements

20.8 Sustainability

- Taking a holistic perspective regarding ecological, societal, and governance aspects (see United Nations sustainable development goals [117])
- Evaluating the use of AI to support reduction, reuse, and recycling of energy or any output; see Appendix P1
- Reflecting on the use of AI to contribute to achieving healthcare as a human right (see Appendix P1); examples include:
  - Identifying patient cohorts with favorable benefit and risk profiles
  - Improving the efficiency of manufacturing processes
  - Streamlining pharmacovigilance processes to identify risks to patient safety
  - Using AI for diagnostics or other medical devices
- Promoting data management standards that allow for reuse of data and models as well as efficient data processing and model engineering; see Appendix M7
- Establishing and adopting governance throughout the AI-enabled computerized system life cycle, adhering to regulatory expectations; see Chapter 3
- Adopting critical thinking and risk-based decision-making to focus efforts and resources; see Appendix M4
- Careful balance of the dynamics of technology evolution and high-quality expectations; see also Appendix P1
- Reflecting on model choices regarding implications on ecological footprint
- Using techniques such as model distillation, i.e. reduction of the model to key capabilities important for the task
- Evaluating the use of cloud resources, particularly when employing hardware such as GPUs, as well as edge computing
- Including sustainability in supplier assessments
- Performing an impact assessment on the use of AI in the workplace, addressing potential fear or stress, but also its contribution to purposeful work

21 Appendix M10 – AI Maturity

21.1 Introduction

A fundamental decision in the design of AI sub-systems is AI Maturity. AI Maturity comprises the dimensions of Autonomy and Adaptiveness; combinations of these dimensions form AI Maturity levels. AI Maturity levels inform a sub-system's control strategy. Decisions on maturity levels should be based on the organization's experience and risk tolerance, the context of use, and risk. These decisions can lead to an AI-enabled computerized system that holds multiple AI sub-systems of varying maturity levels.

21.2 AI Sub-System Autonomy

Autonomy is the degree by which the AI sub-system relies on embedded controls rather than human verification of functions that impact patient safety, product quality, and data integrity. Autonomy is divided into five stages:

- **Autonomy Stage 1:** The AI sub-system runs in parallel to the process and is not used for decision-making. Input and output data of the AI sub-system and its model can be collected.
- **Autonomy Stage 2:** The AI sub-system supports decision-making in a process with an active human decision step for all model outputs.
- **Autonomy Stage 3:** The AI sub-system executes a step or part of a process to make decisions with control and correction by a user. This requires the user to actively monitor, and act as needed to correct the model outputs.
- **Autonomy Stage 4:** The AI sub-system executes a step or part of a process to make decisions and controls itself and signal users to intervene when input data or model output is outside the specified range.
- **Autonomy Stage 5:** The AI sub-system executes a step or part of a process to make decisions and controls and corrects itself.

  Compared to Autonomy Stage 4, in the situation of data input or model output that is out of specification, the sub-system may acquire new data to create new model output until a pre-defined level of certainty is met or eventually signaling users to intervene once a stop criterion is met.

AI sub-systems of Autonomy Stages 2 to 5 may have a direct impact on patient safety, product quality, and data integrity. All human interactions with the system need to be recorded and include the reason for acceptance or corrections of the AI sub-system's model output, as applicable.

Note: An AI sub-system may serve various process steps with differing stages of autonomy, in which case the AI sub-system autonomy is the highest autonomy stage of all such steps.

21.3 AI Sub-System Adaptiveness

Adaptiveness is the AI sub-system's ability to automatically perform updates and thereby automatically self-improve. is divided into six stages:

- **Adaptiveness Stage 0:** The AI sub-system relies on fixed, rule-based, complex algorithms. These sub-systems do not rely on training data, even though the development of the algorithm may exhibit a similar approach as described in Appendix P2. However, changes to the algorithm are manually initiated and implemented.

- **Adaptiveness Stage 1:** The AI sub-system relies on ML and operates as a static system with manual assessment of the need for and initiation of updates. Updates are performed by manually initiated retraining or fine-tuning of the model with new training data sets. Such updates are deployed after manual verification of the new model's fitness for purpose.

- **Adaptiveness Stage 2:** The AI sub-system relies on ML and is operating as a static system, self-assessing and signaling the need for updates to be initiated and verified manually. Such updates are deployed after manual verification of the new model's fitness for purpose.

- **Adaptiveness Stage 3:** The AI sub-system relies on ML and operates as a static system, self-assessing the need for and performing updates to be verified manually.

    - This requires the sub-system's functionality to automatically select data for training or fine-tuning, evaluation and testing purposes, execution of training or fine-tuning processes, and management of model artifacts. The only manual activity during model iterations is verification of the model's fitness for purpose before deployment of the updated model version.

- **Adaptiveness Stage 4:** The AI sub-system relies on ML and operates as a dynamic system, self-assessing the need for, performing, and verifying updates, with manually set objectives.

    - To achieve fully automated learning behavior, in addition to the automation required in Adaptiveness Stage 3, verification of the model's fitness for purpose and deployment needs to be automated. Manual intervention of the learning behavior is only foreseen if the model is out of specification or when changing the objectives.

- **Adaptiveness Stage 5:** The AI sub-system relies on ML and operates as a dynamic system, self-assessing the need for, performing, verifying updates, with self-determined objectives.

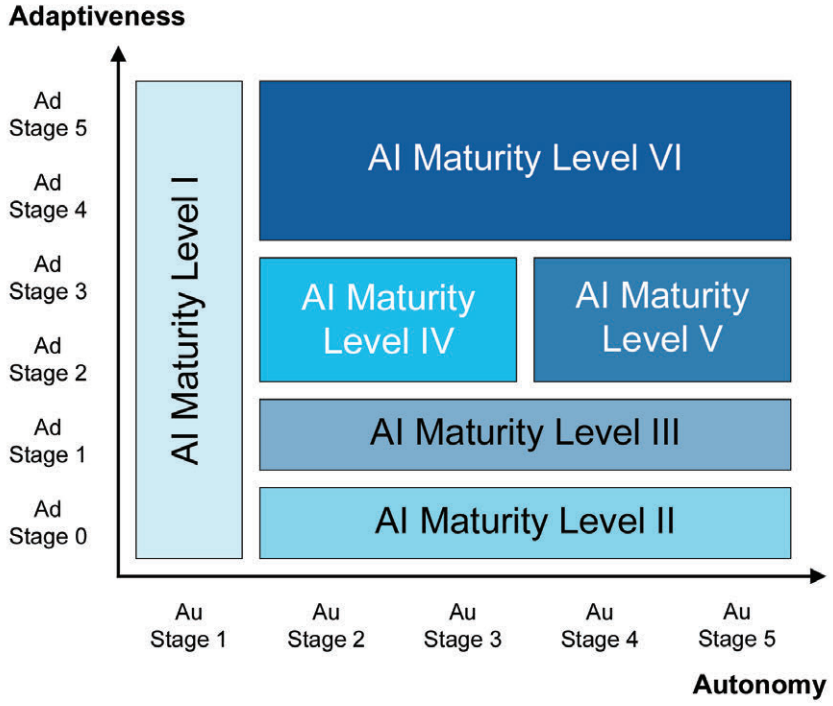
    - Manual intervention of the learning behavior is only foreseen if the model is out of specification.

Note: An AI sub-system may hold various models that may differ in their adaptiveness; in which case the adaptiveness of the AI sub-system is the highest adaptiveness of all such models.

21.4 AI Maturity Levels

Maturity levels are determined in a two-dimensional landscape, based on a combination of autonomy and adaptiveness stages. A typical split, providing six maturity levels, is shown in Figure 21.1, while regulated companies may choose to exclude certain combinations or introduce more granular maturity levels.

Figure 21.1: AI Maturity Landscape Example (adapted [183])



Regulated companies may use these maturity levels to group similar AI sub-systems for deriving control strategies. Detailed controls should be defined following a risk-based approach, see Appendices M3 and M4. Typical control strategies are as follows:

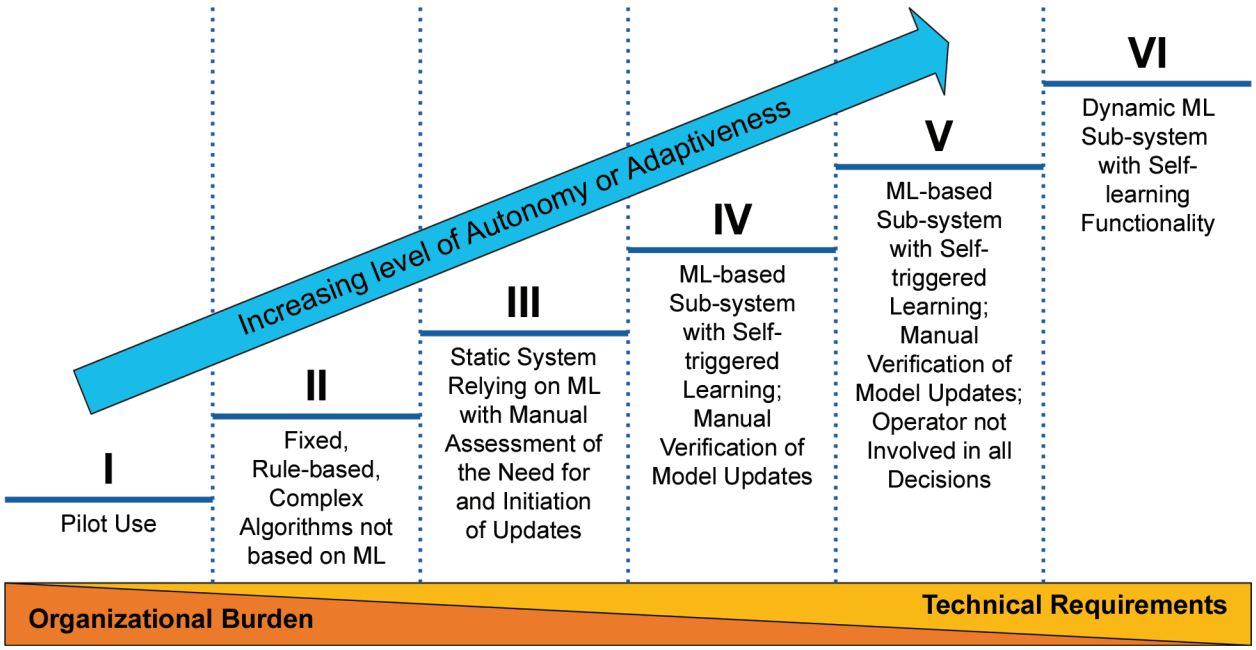
- **AI Maturity Level I:** AI sub-systems are typically used for pilots in a controlled timeframe. Note that they may have an indirect impact on patient safety, product quality, and data integrity.
    - Examples of controls include informing stakeholders that the pilot's results should not be used for decision-making, ensuring that the AI sub-system's computation needs do not negatively impact the performance of the computerized system.
- **AI Maturity Level II:** AI sub-systems rely on fixed, rule-based, complex algorithms and are not based on ML.
    - Example controls for such rule-based algorithms include the selection of a data set that is fit for purpose to evaluate the performance of the rule set in a variety of scenarios in the context of use. Furthermore, acceptance testing should ensure that the algorithm is sufficiently interpretable by end users to base their decisions on the algorithm's results.
- **AI Maturity Level III:** AI sub-systems rely on ML and operate as a static system with manual assessment of the need for and initiation of updates. They are based on training via or fine-tuning with data to establish model artifacts integrated into the system for the generation of outputs.
    - In addition to the controls mentioned in AI Maturity Level II, further controls should be considered, such as verification of the reliability of the sourced model or training algorithms.

- **AI Maturity Level IV:** AI sub-systems exhibit greater adaptiveness, than those in AI Maturity Level III, as varying aspects of the change process are automated. These include signaling the need for, or even performing the retraining or fine-tuning, though involving manual verification of the updated model's fitness for purpose.
    - In addition to the controls outlined in AI Maturity Level III, further controls should be considered, such as verification of the monitoring and trigger mechanisms informing the need for or performing retraining and fine-tuning, and assurance of the data quality used for retraining and fine-tuning.
- **AI Maturity Level V:** AI sub-systems exhibit greater autonomy than those in AI Maturity Level IV. These include capabilities to control or to correct the model's output.
    - In addition to the controls of the preceding maturity levels, further controls should be considered, such as verification of the suitability of controls of the model input and output, as well as the reliability of indications of uncertainty that are used for notification purposes.
- **AI Maturity Level VI:** AI sub-systems exhibit self-learning functionality and thus are dynamic systems.
    - In addition to the controls of the preceding maturity levels, further controls need to be established, for instance verification of the fully automated retraining or fine-tuning process or the suitability of the limitations imposed on the sub-system's self-learning capabilities. Of note, such designs may not be permitted in some regions or GxP areas per regulatory guidance [43].

See Figure 21.2 for an overview of these six AI Maturity levels.

Based on the choice of the Maturity Level and the respective Autonomy and Adaptiveness, trade-offs may arise between the operational burden of controlling the AI-enabled computerized system and its AI sub-systems (more pronounced at lower maturity levels) and the controls that secure an AI sub-system of higher-level maturity.

Figure 21.2: Overview of Example AI Maturity Levels (adapted [183])



22 Appendix M11 – Categories of Software and Hardware

22.1 Introduction

This appendix provides an overview of GAMP software and hardware categories and contextualizes these categories with respect to AI-enabled computerized systems and their components. As stated in *ISPE GAMP 5 (Second Edition), "software and hardware components of a system may be analyzed and categorized in terms of increasing complexity, novelty, and inherent likelihood of residual defects, as a very high-level preliminary risk assessment."* [2]

In addition to GAMP categories, models can be categorized by involvement of data stemming from the regulated company, the supplier, or both. These categories have implications on life cycle activities such as data management, supplier management, change management, knowledge management, and risk management, as discussed throughout this Guide.

Note: Computerized systems are generally made up of a combination of components from different categories; thus, categories should be viewed as a continuum. *"The software category is just one factor in a risk-based approach; the life cycle activities should be scaled based on the overall GxP impact, complexity, and novelty of the system."* [2] See Appendix M3.

"Software categories still bring benefit in deciding the rigor of supplier assessment and also when judging the probability of a failure or defect occurring in a system." [2] Meanwhile, model categories provide an indication of the complexity of models, thus informing life cycle activities.

This appendix is intended to be used in parallel with *ISPE GAMP 5 (Second Edition)* Appendix M4 [2].

22.2 Categories of Software

22.2.1 Category 1 – Infrastructure Software, Tools, and IT Services

"Infrastructure elements link together to form an integrated environment for running and supporting applications and services." These include:

- Established or publicly available layered software such as *"operating systems, database managers, programming languages, middleware, ladder logic interpreters, statistical programming tools, and spreadsheet packages (but not business applications developed using these packages)"*
- Infrastructure software tools such as *"monitoring software, batch job scheduling tools, security software, antivirus, and configuration management tools"*
- *"Software, systems, and tools supporting computerized system life cycle activities and IT and infrastructure processes"* [2]

In the context of AI-enabled computerized systems, various infrastructure elements are used as part of an MLOps architecture, including functionality for data storages, model and artifact management, as well as programming languages and packages used to establish models. See Appendix M8.

"All infrastructure software [tools] should be controlled and managed." [2]. See also *ISPE GAMP Good Practice Guide: IT Infrastructure Control and Compliance (Second Edition)* [111] and *ISPE GAMP Good Practice Guide: Enabling Innovation* [85].

22.2.2 Category 2 – Not Used

Category 2 is not used in *ISPE GAMP 5 (Second Edition)* [2].

22.2.3 Category 3 – Standard System Components

> *"This category includes off-the-shelf components used for business purposes. It includes both those that cannot be configured to conform to business processes and those that offer limited configurations using factory-provided values or ranges (also called parameterization, as may be found in process control systems and simple laboratory devices)."* [2]

Specific to AI, Category 3 systems include those with supplier-provided models that offer no or only limited configuration and adaptiveness to the data of the regulated company. Those models may be embedded in other components (e.g., equipment) or may be provided as solutions (e.g., chatbots with a fixed model).

22.2.4 Category 4 – Configured Components

> *"Configurable software components enable configuration of user-specific business processes into one or more workflows, specific to methods, or products, or processes, etc. This typically involves configuring predefined software modules, and correspondingly there is an increase in the importance of capturing and managing the configuration choices."* [2]

Specific to AI, configured components may offer choices in model configurations or training of models within a set of models included in the configurable component. Deriving a prompt strategy as a configurable item of a product may also constitute a Category 4 software.

22.2.5 Category 5 – Custom Applications and Components

> *"These applications, subsystems, or components are developed to meet the specific requirements of the regulated company. The risk inherent with custom software is high because there is no user experience or system reliability information available…"* [2]

Specific to AI, custom applications include implementation of custom models, training of sourced models based on the regulated company's data sources, and fine-tuning of provided models with own data as an element of a custom application.

22.3 Categories of Hardware

> *"These hardware categories are provided for information only, and do not explicitly require additional documentation."* [2]

22.3.1 Hardware Category 1 – Standard Hardware Components

> *"The majority of the hardware used by regulated companies fall into this category."* [2]

22.3.2 Hardware Category 2 – Custom Built Hardware Components

Custom built hardware components may be needed for specialized AI use cases, such as when high levels of computing power or high data throughput are required. For example, a regulated company may decide to establish custom hardware components to support the training procedures of complex ML models that exceed the capabilities of standard hardware, when sourcing cloud resources is not favored by the regulated company.

22.4 Model Categories

Models can be divided into four categories, based on their involvement with regulated company data, supplier data and data in establishing the model:

- Rule-based, non-ML AI sub-systems
- ML sub-systems trained by the regulated company
- Pre-trained ML sub-system with fine-tuning
- Pre-trained ML sub-system without fine-tuning

Key data and life cycle considerations apply as follows.

22.4.1 Rule-Based, Non-ML AI Model

No data is used for training, although the validation data sets of the regulated company or supplier may be used to evaluate model performance and testing. The regulated company needs to ensure that test data sets used for testing the model or the AI-enabled computerized system are fit for purpose in their context of use. Examples include rule-based knowledge retrieval applications or rule-based visual inspection algorithms.

Changes should be introduced in the same way as for non-AI-enabled computerized systems; verification typically includes testing activities, relying on data that is fit for purpose in the regulated company's context of use.

22.4.2 ML Model Trained with Data from the Regulated Company

All data including training, validation and test data is solely managed by the regulated company. An example is open-source or supplier-provided algorithm that the regulated company uses to train an image classification model.

Changes are initiated by the regulated company without dependency on supplier data.

22.4.3 Pre-Trained ML Model with Fine-Tuning

Data sets are managed by suppliers (pre-trained model) and regulated companies (fine-tuned model) in their respective scope. An example is the use of a pre-trained and fine-tuned neural network for computer vision.

In this category, there are two sources for new model versions:

- The **supplier** manages the pre-trained model, providing iterative versions based on new data or modifications of the algorithm applied to train the pre-trained model.
- The **regulated company** introduces further changes by incrementally fine-tuning the model with the use of own data.

When the pre-trained model is updated, the regulated company needs to consider the implications of a) staying at the current model version (potentially more stable), or b) applying fine-tuning to the new model version (potentially beneficial, though requires greater effort to demonstrate fitness for purpose and avoid unintended consequences).

22.4.4 Pre-Trained ML Sub-System Without Fine-Tuning

Training data is solely managed by the supplier, while validation data may be used for model engineering and test data for model testing activities suitable for the regulated company's context of use, depending on the use case. An example is the use of a LLM as part of a chatbot.

Changes to the model are dependent on activities by the supplier and are considered for use by the regulated company.

23 Appendix S1 – Organizational Changes

23.1 Introduction

Computerized systems are affected by organizational changes. Therefore, regulated companies should consider the impact of organizational change on AI-enabled computerized systems to maintain their effectiveness and safety. Compared to other implications of organizational change, the impact on data, models, knowledge, and governance practices should be considered.

This appendix highlights aspects that are relevant to supporting the safe and effective use of AI-enabled computerized systems in the context of organizational change. This guidance augments information provided in *ISPE GAMP 5 (Second Edition)* Appendix S6 – Organizational Change [2], which includes initiators for change, scope and impact of change, organizational factors, outsourcing, loss of a supplier, risk assessment of organizational change, and affected stakeholders.

23.2 General Considerations

Key aspects that regulated companies should consider for AI-enabled computerized systems in the context of organizational changes include:

- Data and model management and governance are of high importance to maintaining control and effectiveness of AI-enabled computerized systems across the organization. During organizational change, they should aim for harmonized data and model management and governance practices to allow for effective use of data and models. For instance, a common set of policies and practices should be established in the event of mergers. See Appendix M7.

- AI-enabled computerized systems may exhibit a high level of complexity, including various contributions from suppliers. Maintaining a state of control for such systems requires deep knowledge. When regulated companies undergo change, they should ensure that sufficient expertise is maintained to support complex systems; of particular interest are new roles and expertise required in the context of AI-enabled computerized systems.

  In addition, knowledge management requires openness, transparency, and collaboration, especially when reflecting on possible weaknesses of data models. This again requires strong culture and leadership, especially in a situation where organizational change may lead to general uncertainty across stakeholders.

  Further information is provided in Appendices M2 and M5.

- In the case of mergers, the increased availability of data and models may provide opportunities to further improve AI-enabled computerized systems or equip existing computerized systems with capabilities that have been proven safe and effective in different settings. However, regulated companies need to carefully assess the suitability of an approach before transfer and any possible restrictions, such as use of sensitive personal data. At the same time, retention periods for data and models should be considered, which may require data and model archives to capture the exact system outline at the time of operation.

  Further information is provided in Appendices M7 and P1.

- In case of mergers, systems and technologies forming the respective IT infrastructure landscape should be assessed, in order to derive a target operating model and a migration strategy, if applicable. Multiple systems for the management of data and models may lead to incompatibility and increased integration complexity.

  Migrating data and models requires compromises in data storage and model repositories, leading to changes in AI-enabled computerized systems. These should be managed according to accepted change management processes.

  See also Appendix M8 for elements of AI-specific IT infrastructure.

24 Appendix S2 – AI Adoption and Challenges

24.1 Introduction

Regulated companies require a wide range of expertise to achieve safe and effective AI-enabled computerized systems, in support of innovation. While aiming for high-quality AI-enabled computerized systems as a regulatory obligation, they need to manage various challenges. As outlined throughout this Guide, collaboration is a key element to success in the adoption of AI, which requires planning and consideration.

This appendix provides an organizational perspective on achieving adoption of AI, covering AI-enabled computerized systems generally.

24.2 General Considerations

Management generally aims for efficient deployment of resources, achievement of business objectives and high quality, and managing variability of workload to achieve the desired utilization of resources.

In the context of AI-enabled computerized systems, management should also aim for the following:

- **Managing the gradient of building knowledge and understanding:** AI projects are often characterized by incremental building of knowledge regarding how the interplay of models and data can support a GxP regulated process effectively and in compliance with applicable regulations.

  Agile concepts can support this process in day-to-day operations of such teams, allowing them to set goals within a limited time frame and then re-navigate. Refer to *ISPE GAMP 5 (Second Edition)* [2] for further guidance on agile principles.

- **Fostering a collaborative culture:** While teams may already possess a variety of expertise and skills, additional experts may be required for consultation or specialized support.

  Management is advised to foster an open environment, where stakeholders may be invited to support initiatives while also managing their involvement in such projects in balance with their day-to-day activities. See Chapter 6 Regulated Company Activities.

- **Aiming for business objectives while promoting high quality:** With substantial uncertainty being managed throughout the life cycle of AI-enabled computerized systems, decision-making should be based on risk with application of critical thinking; see Appendix M3 and M4. Considering limited resources, thoughtful decisions are required particularly for the first introduction and subsequent major revisions of an AI-enabled computerized system.

  A balance needs to be met to gain initial benefits in a shorter time frame, while keeping further potential within reach. Regulated companies may establish a roadmap, underpinned by a comprehensive rationale. Such rationale should be developed and supported by a wide range of stakeholders, requiring effective communication and moderation. See Appendix P1 on conceptualization and scope definition for AI use.

- **Leveraging supplier involvement sustainably:** Suppliers may hold relevant knowledge about the implementation and use of AI, though they may have varying levels of understanding about the regulated company's specific conditions and goals. A high degree of transparency is required, building collaborative approaches to allow for the transfer of know-how in both directions.

24.3 Challenges

The adoption of AI poses many challenges to regulated companies. These challenges span the entire life cycle and comprise various areas, including:

- Regulatory aspects
- Innovation and technology related challenges
- Organizational challenges
- QRM aspects and scaling of life cycle activities

Understanding typical challenges can help to prepare for potential obstacles and thus support successful implementation of safe and effective AI-enabled computerized systems.

24.3.1 Knowledge Management

Various challenges related to knowledge management [82] include:

- Building new skills, acquiring talent, developing AI literacy and skills among an established workforce, and offering new career paths to foster deep AI expertise
- Managing the heterogeneity of language, particularly terminology, and the heterogeneity of information collected
- Linkage of knowledge management and risk management when using AI, including translating knowledge into meaningful risk-based decisions and leveraging risk management activities to build knowledge
- Managing innovation, keeping up with the rate of change, and managing employee uncertainty by quickly providing them with new skills
- Retention of knowledge and skills; although some roles may require less once AI is implemented, existing knowledge and skills bases may already be scarce and affected by fragmentation

Guidance on how to augment knowledge management systems is provided in Appendix M5.

24.3.2 Regulatory Expectations

At the time of writing, regulatory agencies are establishing regulations and guidance, responding to sponsor submissions, and developing further guidance on the use of AI in regulated areas. A list of relevant Statutes, Regulations, Standards, and Guidance documents is provided in Appendix S7.

This evolving regulatory landscape can pose challenges, including the following:

- Most of the advanced guidance is available for AI in medical devices, with only high-level guidance for other GxP areas (at the time of writing)
- Application of guidance to a specific use case can be challenging, as it does not reflect nuances of GxP areas
- The combination of regulation and guidance specific to life sciences and cross-sectoral regulation like the EU AI Act needs to be considered [24]
- Developments in the wider regulatory landscape are relevant, such as the Machinery Regulation (EU) 2023/1230 [118] and its considerations on AI in manufacturing as well as its link to other regulation such as the EU AI Act [24]

- Monitoring of evolving guidance requires resources with deep expertise in life sciences and regulatory areas
- Regulatory guidance by different health authorities often overlap; however, regulated companies should address potential regional nuances
- Technology advancements may require further regulation, leading to continued dynamics in the regulatory landscape
- As health authorities encourage early interactions (e.g., EMA reflection paper [43] and FDA discussion papers [44]), it is important to strike the right balance between engaging in early discussions and ensuring sufficient maturity in understanding and planning to facilitate a constructive exchange.

24.3.3 Supplier Relationships

With the emergence of AI, enhancements to existing computerized systems, and the development of new AI-enabled software products, regulated companies may face challenges in managing their supplier relationships, including:

- Thorough assessment of their suppliers, while experiencing difficulties in the complexity of new products and services, as well as data management and model engineering practices
- Establishing a common language when interacting with their suppliers in the context of AI
- Assessing the fitness for purpose of data provided by their suppliers to support AI use cases
- Aligning and achieving the right balance between protection of IP and transparency regarding AI-enabled software products
- Aligning rules on the use of data for secondary purposes

This Guide is designed to offer guidance to regulated companies in appropriately managing suppliers in the context of AI-enabled computerized systems; see Chapters 6 and 7 and Appendix M2.

24.3.4 Speed of Technology Innovation

Challenges related to keeping up with the speed of technology include:

- Experiencing a growing variety of models and data, providing more options to choose from when selecting an exact model that is fit for purpose
- Designing systems such that they are capable of embodying changes in the future and thus can accommodate new data and technical advancements
- Balancing facilitation of continual improvement [82] with the need for safety and robustness of computerized systems

To overcome this challenge, this Guide reflects comprehensive decision-making along a tailored life cycle, see Chapters 3 and 4 and Appendix P1 for further details.

24.3.5 Risk-Based Decision-Making

Specifically for AI-enabled computerized systems, even decisions made in early phases can introduce risks during operation. Therefore, conventional risk management activities need to account for additional aspects in the use of potentially complex models and their interplay with data and end users.

A common challenge is the lack of standardized frameworks for managing AI-related risks.

This Guide covers AI-specific aspects of QRM in Chapter 5 and Appendix M3, including a risk-based approach to right-size the rigor and effort for testing.

24.3.6 Fit for Purpose Data

It is important to identify data that is fit for purpose in the context of use. Challenges associated with the need for data that is fit for purpose include:

- Hurdles in defining and obtaining adequately representative data sets due to the innate difficulty in understanding how relevant scenarios can be reflected by data in a complex domain
- Assessment of the data's fitness for purpose requiring significant resources, including both technical disciplines and domain expertise
- Data still being captured non-digitally in many scenarios, restricting effective utilization for the development and use of AI-enabled computerized systems
- Shortcomings in collected data, such as bias, requiring thoughtful decision-making during all life cycle and validation activities

Guidance is provided in Appendices M6 and M7 on overcoming these data challenges.

24.3.7 Cybersecurity Management

Security risks for AI-enabled computerized systems vary from traditional intrusion attacks to AI-specific cybersecurity threats.

For instance, malevolent actors may introduce slight changes to model inputs, which can be difficult for humans to identify but may have a substantial effect on the model output. This example demonstrates the challenges in the interplay of complex models operating on data and underpins the need for robust cybersecurity management practices.

Guidance on AI-specific cybersecurity is in Appendix S5.

24.3.8 Legal Considerations

Legal considerations are relevant in life sciences industries due to the potential impact of decisions and processes on patients and users. These include GxP regulations, as well as generally applicable statutes regarding IP rights [119] or privacy of personal data ([25] for example) as well as cross-sectoral regulation on the use of AI.

Regulated companies may be held accountable for damage or harm to patients, individuals, or other organizations. Further challenges arise from the dynamic legal and regulatory landscape surrounding the use of AI. These dynamics may lead to new requirements that need to be addressed during the development of an AI-enabled computerized systems, or even for those systems already in use.

This Guide does not provide legal advice; it promotes practices and frameworks that support achieving legal compliance and leveraging process understanding and critical thinking.

25 Appendix S3 – Machine Learning (ML) Fundamentals

25.1 Introduction

This appendix provides an overview of fundamental technical details in the context of ML. It is intended for stakeholders interested in conceptual ideas and relationships between learning strategies and model types.

25.2 Learning Strategies

Learning strategies refer to different strategies by which an ML model derives patterns from data and thus can be used to train a model. This appendix provides an overview of typical learning strategies. The specific choice depends on the amount or nature of the training data and the problem being addressed. See Appendices P1 and P2.

25.2.1 Supervised Learning

Supervised learning is a learning strategy used when a full set of labeled data is available to train a model. Each input example is labeled as ground truth, which represents the desired model output. Examples of data labeling include:

- Assigning labels to images to identify objects, actions, or scenes
- Tagging text to indicate sentiments, topics, or entities such as names, locations, and dates
- Transcribing audio files to indicate what is being spoken or to identify speaker emotions or characteristics

Supervised learning is used where a set of reference ground truth is available for the model input and where the output is well defined.[22]

Common use cases include classification tasks (where each element of the input data is identified as one or more of a predetermined set of classes) or regression tasks (where a specific target variable is predicted).

25.2.2 Unsupervised Learning

Unsupervised learning describes when a model is trained using unlabeled data, i.e., without a tagged or annotated ground truth. The goal is to *"uncover hidden patterns or structure within the data"* [120] without any specific target output.

Unsupervised learning is used in cases where fully labeled data sets are not easily available or when the desired output of the ML model is not known at training time. Unsupervised learning utilizes techniques that can self-identify patterns in the data, effectively categorizing inputs into a set of derived classes or mapped onto one or more latent variables.

Examples of use include:

- Unsupervised classification, which employs models that accept a collection of multi-variate input data and derive its own set of classes into which the data is partitioned

[22] Supervised learning is described in the FDA Digital Health and Artificial Intelligence Glossary – Education Resource as *"ML algorithms where labeled data is provided, and algorithms are trained using the labeled data. Labeling or annotation is the process of attaching descriptive information to data. Data itself is unchanged in the annotation process."* [120] Self-supervised learning is described as *"ML algorithms that generate their own labels from the available unlabeled data."* [120] A typical use case is the training of LLM, where self-supervised learning is constructed by prediction of the next word, only a fraction of the otherwise unlabeled text is provided, and the following word is considered as ground truth.

- Identification of patterns in complex or noisy data
- Dimensionality reduction, which maps complex input variables into a smaller, more manageable set of variables
- Anomaly detection, which identifies outliers according to patterns that deviate from known normal operation

These use cases render unsupervised learning commonly used in early research, though applications also exist in stricter environments.

Many unsupervised learning algorithms benefit from the addition of regularization parameters or strategies during training and can require hyperparameters, allowing for iterative experimentation.

25.2.3 Semi-Supervised Learning

Semi-supervised learning falls between supervised and unsupervised learning. The ML model is trained using a combination of labeled and unlabeled data, with the goal of improving performance on the task while minimizing the amount of labeled data needed.

Example use cases of semi-supervised learning include:

- *"Situations where obtaining a sufficient amount of labeled data is prohibitively difficult or expensive*" [121]
- The use of models to generate a richer set of data that can then be added to the manual data to train a secondary ML algorithm. As an illustrative example, medical image object segmentation benefits from semi-supervised learning: The manual ground truth consists of images annotated by clinical experts to identify anatomy or lesions. Generative AI techniques such as GANs can then be used to create new plausible examples of images and annotations. Once verified, these generated examples can be added to the case data to train a model.

25.2.4 Reinforcement Learning

Reinforcement learning is a means by which the model learns by interacting with the environment and receiving feedback in the form of *"rewards or penalties"* [120] for certain actions. This feedback makes it more likely that the system repeats the correct decisions and is less likely to make incorrect decisions in future similar scenarios.

Reinforcement learning can continually adapt and improve the performance of the model to achieve a known, desired goal. Given a particular state and set of possible actions, the model learns which actions are likely to lead to the desired outcome by testing several actions and rewarding the one that leads to the outcome, thus making the choice of that action more probable in future, similar states.

For example, a robotic manufacturing system may be faced with multiple paths to transport an item from one location to another. Reinforcement learning encourages the model to try multiple paths given similar situations and reinforce the path that is faster, more economical, or less likely to result in failure.

Reinforcement learning is suitable in cases where immediate feedback loops are available in the process, and there are multiple choices to achieve an outcome that can be closely evaluated by a minimization function fitting the use case's purpose.

25.3 Model Types

This section provides an overview of various model types typically used. While it is not a complete list, considering the dynamics in data science, it serves to illustrate the learning strategies discussed in Section 25.2.

25.3.1 Classification Models

Classification models take model input and predict a label. These models predict a probability, so that rules (such as highest probability among all classes) can be applied to assign a predicted class.

In binary classification there are only two outputs (e.g., Yes, No or A, B), while in multivariate classification the model produces multiple outputs to indicate probabilities for each class.

25.3.2 Cluster Analysis and Models

Cluster analysis and models group observations that share similarities, thus deriving classes. Clusters can be equipped with a semantic description. Cluster analysis does not require any training data; therefore, clustering is often used in data preprocessing to structure data. Additionally, clustering is often used in aspects of R&D or clinical trials as they can categorize many different types of diagnostics or other qualities.

25.3.3 Regression Models

Regression is a statistical method in which a regression function is determined from a data set, which functionally represents the relationship between the independent variables selected as model input and the dependent variable selected as model output. Regression models construct a function that calculates the dependent variable's value based on the combination of independent variable values.

A technique that is regularly applied to regression models is model calibration, while it can also be used for other model types. Model calibration is the *"process of adjusting predicted probabilities generated by an ML model to ensure that they accurately reflect the observed frequencies of events or outcomes in the real world."* [120] Various techniques can be used, including global calibration or calibration applied to classes of model input and/or model output. The resulting relationship between the input and final output should be assessed regarding potential implausible behavior, like larger jumps or non-monotone behavior when monotonicity is expected.

25.3.4 Decision Trees and Random Forests

A decision tree uses a tree-like structure to model decisions. It consists of a base that accepts a multi-variable data set, with branches leading to different decisions based on the value of a specific or small grouping of variables.

For example, a binary tree might involve branching right if a value in the input data exceeds a threshold at the node and left if it does not. This process is repeated until a tree branch is reached to deliver the model output. Decision trees can have more complex multiple branch structures.

Decision trees are effective at assisting with inferences in decision-making and dealing with both non-numerical and numerical data to support the process's robustness.

Random forests, an extension of the decision tree, are an example of an ensemble learning method which generally relies on a combination of various base models. It consists of several decision trees; each tree in the forest determines its class, while the final model output is based on the class with the most votes. With larger combinations of multiple decision trees, the results have reduced variance and typically provide more robust model performance.

25.3.5 Support Vector Machines (SVM)

Support Vector Machines can be used for classification tasks or prediction of target variables. They are effective for cases of non-linear relationships between model input and model output.

SVMs take the most effective separating detail (often called “finding the optimal hyperplane”) in the featured data and separate them into groups based on those data points. [153] The hyperplane operates as a line in two dimensions or plane in more dimensions, separating one group from another and acting as that most effective separating detail.

25.3.6 Principal Component Analysis (PCA)

PCA is a statistical procedure by which base components in a data set are identified and organized by weighted importance.

A common use case includes limiting the number of model input by selecting the most relevant components, therefore decreasing computing cost and simplifying complex data sets. PCA can also be used to identify key features of the data.

25.3.7 K-Nearest Neighbor (KNN)

KNN is a classification-based algorithm that assigns data points where an instance is assigned to a class based on the values or “vote” of its neighbors and similarity to those neighbors.

KNNs are efficient to implement and facilitate ease of visualization of data. With this simplicity comes easy explainability and the ability to handle numerical and categorical data, providing flexible options. KNN could be used for tasks such as pattern recognition or recommendations.

25.3.8 Naïve Bayes (NB) Algorithm

NB is a classification algorithm based on the Bayes’ theorem of conditional probability, which states that the likelihood of an outcome accruing can be based on the previous outcomes of similar circumstances.

The model operates on two assumptions: every pair of data points is independent of one another and the weights of each pairing are equal (i.e., no pair impacts decision-making more than the other). The model calculates the probability distributions of events A or B, in the example case of binary classification, under these assumptions.

NB provides a strong yet simple means for grouping data and potentially making decisions based off a data set. Also, NB could be used in various scenarios such as classifying text or detecting communication in spam.

25.3.9 Neural Networks

A neural network is a network of connected artificial neurons, *“inspired by the structure of the human brain”* [120]. Each neuron consists of certain functions to convert an input to an output, eventually deriving the model output. The number of nodes and layers define the complexity of the neural network. These methods emulate how biological neurons work together to identify scenarios and reach a conclusion.

In general, a neural network has an input and output layer and may exhibit hidden layers in between. A network with many hidden layers is called a deep neural network and is the origin of the term deep learning. It accepts multiple inputs, each of which are multiplied by their weights and then summed. The resultant value is fed into an activation equation that determines a binary yes-or-no response (“activation”). Activation and output values can be fed as inputs to neurons in the next layer, essentially passing data from layer to another.

Various types of neural networks exist, and they can be distinguished according to their structure. A selection of example structures is provided in the following subsections.

25.3.9.1 Feed Forward Neural Network

Feed forward is the simplest form of neural networks where information is fed straight from the input, through the hidden layers to the output. Feed forward neural networks are effective for use in pattern recognition and tasks related to regression and classification.

25.3.9.2 Multilayer Perceptron (MLP) Neural Network

A MLP neural network stems from the feed forward neural network and is a model consisting of multiple layers of neurons that enable the ability to solve complex problems by learning non-linear relationships in data sets. Examples of MLP include image processing and NLP.

25.3.9.3 Convolutional Neural Network (CNN)

CNN is primarily used for image processing. Generally, it accepts image pixel values as model input and is organized to apply a series of filters to the data transverses through the network. Filters are typically applied multiple times as a moving window across the image or specific sections of the image.

As suggested in the name, CNN is analogous to performing a series of convolution filter operations over an image. The network may also be able to employ multiple filters tailored to find a specific quality of the image that is being processed. For further information, see the FDA Digital Health and Artificial Intelligence Glossary – Educational Resource website [120].

25.3.9.4 Recurrent Neural Network (RNN)

In addition to the feed forward structure, RNNs contain one or more feedback loops and are effective in processing time-series data sets. The nodes in this model are often in a specific sequence, meaning one node depends on the previous node. The RNN then separates itself from the base feed forward neural network. This architecture enables the network to contain a hidden state, which allows them to retain information from the previous time steps, thus allowing them to capture dependency in a time series of data and learning to predict the likely next occurrence in a sequence of data provided.

A Hopfield Network (HN) is an example of an RNN. In an HN, every node is connected to one another with connections carrying a specific strength and weight. HNs could be used for tasks like image processing.

Long Short-Term Memory (LSTM) neural networks overcome an issue found in RNN called “the vanishing gradient” that occurs during training when the changes requested of the network weights become vanishingly small. LSTMs overcome this by organizing its input neurons and output neurons to make determinations on how much is learned from the previous node, allowing for the model to forget certain details from the previous layer and subsequently produce a more stable prediction than a basic RNN.

25.3.9.5 Transformer Networks

The transformer network architecture, proposed in a 2017 paper [122], forms the basis of many conversational-like generative models. A transformer consists of an encoder and decoder:

- The **encoder** takes the inputs and encodes them in a fixed-length representation.
- A **decoder** uses this initial hidden state to generate the first token of the output sequence.

It then generates subsequent tokens by updating its internal state based on the output thus far, until an end-of-sequence token is generated or a maximum sequence length is reached. Variations of these transformer networks consist of a decoder only (e.g., some Generative Pre-Trained Transformer models (GPT)), an encoder only (e.g., Bidirectional Encoder Representations from Transformers (BERT)), and both decoder and encoder (e.g., Bidirectional and Auto-Regressive Transformers (BART)).

26 Appendix S4 – Explainable AI (XAI)

26.1 Introduction

AI sub-systems have varying degrees of complexity, as described in Appendix S3. As complexity increases, it typically becomes more challenging to understand and interpret model output for both technical and non-technical stakeholders. Thus, there is a need for supporting methods to allow for interpretation of model output. XAI methods can support achieving this goal. This is particularly important when considering aspects of human-AI interaction, such as verification or *ex post* assessments of model output.

Recommended measures for regulated companies include:

- Developing an understanding of the need for XAI methods
- Identifying situations where XAI methods can be applied
- Selecting adequate XAI methods that meet expectations of end users
- Being aware of the use of XAI methods for development purposes
- Establishing effective communication among stakeholders to derive suitable and effective XAI methods

XAI should offer explanation, be meaningful, provide explanation accuracy, and adhere to knowledge limits [123]. To ensure these principles are maintained throughout the AI-enabled system's life cycle, close collaboration among members of the project team and further stakeholder ecosystem is required (see Section 26.3). To facilitate this collaborative effort, the organization needs to establish AI literacy (see Appendix M5).

Via this collaborative effort, decisions to derive XAI should be made on an informed basis, taking advantage of insights throughout the life cycle such as the results from the PoC, from model engineering and evaluation activities, and ongoing monitoring including user feedback; see Appendices P1, P2, and P3.

XAI can help to maintain ethical principles, such as safety or human autonomy, contributing to trustworthy AI, see Appendix M9. Regulated companies are encouraged to include considerations on XAI in their AI Governance, thus guiding teams and stakeholders in their activities and contributions.

26.2 XAI Principles

Explainability is the degree to which a basis for a decision or action can be explained or how an output or result was reached, in a way that a person can understand [33].

Applying the concept of explainability to AI, XAI satisfies four principles, according to NIST [123]:

- *"**Explanation:** A system delivers or contains accompanying evidence or reason(s) for outputs and/or processes.*
- ***Meaningful:** A system provides explanations that are understandable to the intended [users].*
- ***Explanation accuracy:** An explanation correctly reflects the reason for generating the output and/or accurately reflects the system's process.*

- ***Knowledge limits:*** *A system only operates under conditions for which it was designed and when it has sufficient confidence in its output."*

Explainability supports interpretability of the system, which is the *"ability to contextualize a model's output in a manner that relates it to the system's designed functional purpose, and the goals, values, and preferences of end users."* [124]

26.3 Stakeholder Contributions to XAI

A collaborative environment in which stakeholders possess a clear understanding of their contributions supports the achievement of XAI. While typical roles, responsibilities, and organizational units may differ between organizations and roles may also be combined, key contributions to achieve XAI include:

- **Process owners:** Ensuring that the business process is accurately reflected in the system and supported by XAI methods so that it meets its objective while aligning with regulatory standards and policies of the company.
- **System owners:** Ensuring that the system provides the intended functionality in a robust manner. They provide input for the choice of XAI methods to achieve feasibility within the system environment.
- **End users:** Expressing their needs to achieve interpretable use of AI within the context of use and can help identify model failure modes based on actual use. End users include business users, patients and consumers, and healthcare professionals.
- **Quality unit:** Supporting alignment with regulations and internal policies or procedures. They ensure that information and records reflect the decision-making of the approach to achieve XAI; they provide oversight on adherence to XAI principles.
- **Business analysts:** Working with data scientists or ML Architects to identify a model(s) that meets business needs. This includes choices between explainable models, and the choice of additional methods to explain the model's results. They work closely together with domain experts who hold a broad understanding of the field where AI is being deployed.
- **ML architects:** Supporting XAI through appropriate design of the system and model architecture, in collaboration with ML and AI engineers and data scientists to include insights from evaluation of data and models.
- **Data scientists, data engineers, AI engineers, ML engineers, and software engineers:** Supporting XAI by implementing AI sub-systems that are consistent with user requirements when embedded in the AI-enabled system and utilize XAI methods described below.

26.4 XAI Methods

Regulated companies should consider the use of XAI methods to describe the "why" behind model outputs in support of achieving XAI principles. These methods can provide a variety of information, including feature summary statistics and visualizations, model parameters, and interpretable model approximations.

The suitability of XAI methods is dependent on the context of use, depending, for example, on the data modality of model inputs and outputs, the needs of end users, and the complexity of the AI sub-system's objective. Infrastructure may need to be considered as well, as XAI methods can be resource intensive.

XAI methods may be model-agnostic or model-specific. Model-agnostic XAI methods such as Local Interpretable Model-agnostic Explanations (LIME) (described in Section 26.7) or SHapley Additive exPlanations (SHAP) offer the following advantages because of their versatility:

- User interfaces leveraging model explanations can be conceptualized independently of the model.
- Data scientists can apply these methods for a wider range of models under consideration in the project phase to gain better understanding of the relation of data inputs and model outputs.

However, if model-agnostic methods do not achieve sufficient explainability, then model-specific XAI methods should be considered. For example, activation heat maps[23] can be applied to interpret a neural network's decision-making process.

26.5 Use of Model Features for XAI

Model features are the individual measurable properties of model input, from which a model has derived patterns to make predictions.

Feature statistics can be used as an XAI method. While some feature statistics produce a singular value indicating the importance of a feature, others offer feature correlation data, such as the strength of interactions between pairs of features. Feature correlations are represented by individual numbers for each pair.

A graphical representation and statistical summaries of model features support gaining insights from feature statistics. Some summaries gain their significance primarily through visual representation, such as plotting a curve to depict the relationship between the model's feature and the model's predicted outcome.

26.6 XAI Methods Based on Model Anatomy

Certain models have anatomies that can be used as XAI methods. For example, weight magnitudes multiplied by the standard deviation of the respective feature provide a relative feature importance for linear models. Another example includes tree structures, such as features and split thresholds, in decision trees.

26.7 Use of Proxy Models for XAI

To decipher the mechanisms of models with high complexity, one method is to approximate them with a less complex proxy model that is easier to explain. LIME is one approach for these proxy models. LIME operates by examining how model outputs change based on variations of the input data. These variations create a new data set comprising perturbed input data along with their model output. Applying a LIME approach, a simpler, interpretable model is trained on this new data set, with a weighting of how close these new pairs of perturbed input data and model output are to the specific data point being analyzed. This simpler model is called a proxy model, as it mimics the behavior of the original, more complex model. For example, the proxy model could be a decision tree. Subsequent interpretation of these models typically involves examining the proxy model's anatomy or statistical summaries of features to explain the original, more complex model.

[23] Activation heat maps are visualization tools used to identify the specific areas of an input image that a neural network uses to recognize a particular data type within the image.

27 Appendix S5 – Cybersecurity Challenges

27.1 Introduction

Cybersecurity is a key concern in the use of computerized systems since cybersecurity attacks can have an impact on patient safety, product quality, and data integrity. While cybersecurity threats are generally relevant to the operation of any computerized system, there are additional aspects that warrant further consideration in the context of AI-enabled computerized systems, including:

- The complexity of model architecture exhibits new attack vectors.
- The reliance on data may lead to unreliable model outputs when compromised.
- Attacks may have impacts beyond the directly affected model output, since ongoing monitoring may be ineffective or future development may be impacted.

Regulated companies should maintain inherent functional security through processes and policies that promote safeguarding patient safety, product quality, and data integrity throughout the life cycle. Regulated companies should safeguard their models from theft, tampering, and illegal access [114]. The specific context of use, the model and system architecture, and humans that interact with the model need to be considered. Controls should be developed based on an understanding of risk, and effectiveness of controls should be verified. See Appendices M3 and P2.

AI can also be used by attackers; this affects cybersecurity strategies specific to AI-enabled computerized systems and the general cybersecurity posture of organizations.

This appendix refers to relevant regulatory frameworks and standards including ISO/IEC 42001 [5] as well as the EU Cyber Resilience Act [125], ISO 27563 [126], EU GDPR [25], ISO 27001 [59], SOC 1/SOC 2/SOC [66], NIST [67], and FDA Pre and Post-Market Cybersecurity guidelines [127, 128] as it examines the direct and indirect cybersecurity concerns related to AI-enabled computerized systems. ISO/IEC 42001 [5], which focuses on AI [Quality] Management Systems, places an emphasis on the necessity of protecting computerized systems generally, and their embedded AI sub-systems and models, at every stage of development (see ISO/IEC 42001, B.6.1.3 Processes for responsible design and development of AI systems).

SOC 2 [129] encourages organizations to establish controls to maintain the security and integrity of computerized systems, hence their AI sub-systems and their embedded models, in accordance with SOC 2 Trust Services Criteria, paying special attention to security, availability, and confidentiality principles (SOC 2, TSC CC6.1).

Further details on cybersecurity considerations are in *ISPE GAMP 5 (Second Edition),* including Appendix O11 [2].

This appendix primarily addresses regulated companies' concerns. Suppliers providing software products may leverage this guidance for possible expectations from regulated companies on their cybersecurity standards.

27.2 Data Protection, Privacy, and Confidential Information

Regulated companies should understand the AI-enabled computerized system, and the use of data as part of the system, in the context of use and within the regulatory jurisdiction it is intended to serve.

This is particularly relevant for personal data, with applicable data privacy regulations such as the EU GDPR [25], while a range of maturity stages and enforcement is present globally at the time of writing. Current privacy regulations might not directly reference AI, though regulated companies should design AI-enabled computerized systems with compliance in mind. Similarly, they should consider the protection of confidential corporate information. See EMA reflection paper, 2.7. Integrity aspects and data protection [43].

Frameworks such as ISO 27001 [59] and HITRUST CSF [130] can be useful in the protection of data privacy and confidential information, in establishing awareness of cybersecurity risks and their impact on personal data, and in suggesting organizational and technical controls:

- ISO 27001 may support establishing strong measures to safeguard the availability, confidentiality, and integrity of data used in AI-enabled computerized systems, see [59] Annex A.8.

- HITRUST CSF provides a thorough method for handling privacy and information security threats for health centric use cases. Special consideration should be given to the CSF's privacy and data protection controls for AI-enabled computerized systems found in HITRUST AI Risk Management Assessment, which is harmonized with ISO/IEC 23894 [8] and NIST AI RMF [93].

27.3 Activities of Malevolent Actors

Models rely on high-quality data for various purposes, including model engineering, use during operation, and ongoing monitoring. Malevolent actors targeting data sources to the AI-enabled computerized system can significantly compromise model behavior and the quality of the data. See considerations on problem and incident management in Appendix P3.

The impact of such activities also depends on the design of the system. For example, a malevolent actor's activity may have direct impact on the model in the case of dynamic systems, exhibiting risks to use of all future model output in the process.

To manage such risks arising from malevolent actors, the FDA Post-Market Cybersecurity guideline [128] places a strong emphasis on the necessity of ongoing risk analysis and surveillance, in the context of medical devices. Regulated companies may leverage these concepts to establish a thorough cybersecurity management strategy, including active monitoring of AI-enabled computerized systems.

Depending on the system exposure and level of risk, employing a Red Team[24] when testing the AI-enabled computerized system is good practice.

27.4 Software Supply Chains

Regulated companies should be aware of potentially complex supply chains in the context of AI-enabled computerized systems, including the use of open-source software, that give rise to unique security concerns.

ISO 27001 Annex A.15 [59] addresses supply chain security and emphasizes the need for supplier monitoring and supplier agreements, relying on trustful relationships to maintain high security standards throughout the supply chain. See Chapter 6 and Appendix M2.

HITRUST CSF [130], HITRUST CSF v9.4, Control Category 09, provides an overview of controls for managing third-party risks for computerized systems that rely on external data sources or services.

The concept of Software or Cyber Bill of Materials (S/CBOM), e.g., for both medical devices and US government supportive systems, can also be helpful in other safety-critical settings. A properly managed SBOM in tandem with alignment with established security frameworks can further strengthen the integrity of the software supply chain.

[24] A red team tests security features as if they were a malevolent actor to gain insights into the robustness of the system and its controls.

27.5 Business Continuity

Implications of cybersecurity attacks, for example on data or models, may lead to complex challenges in determining the best possible path to re-establish operations.

For this reason, business continuity planning should begin early in the development of the AI-enabled computerized system and its AI sub-systems. The potential impact of a partial or full loss of access to an AI system should be considered an implication of the attack.

Leveraging critical thinking, regulated companies should evaluate the impact of the system outages on patient safety, product quality, and data integrity.

Further information on business continuity planning is in Appendix P3.

27.6 Moral and Ethical Considerations

Ethical aspects should be considered with respect to the potential wider implications of a security breach. This is of elevated importance in AI-enabled computerized systems with direct linkage to patients, such as for medical devices or management of patient data in clinical trials. While disclosure of information in these areas may have severe consequences for an individual person, ethical aspects may also be relevant in other GxP areas. For example, a security breach in the context of pharmacovigilance, which alters substantial parts of a safety database, may have ethical implications as the signals on patient safety may be diluted.

Therefore, regulated companies should address ethical concerns regarding AI-enabled computerized systems to maintain security and confidence. ISO/IEC 42001 [5] promotes taking ethical considerations into account when managing an AI-enabled computerized system. See Appendix M9.

27.7 Examples of Cybersecurity Threats

The following scenarios illustrate the challenges and threats posed by adversarial AI scenarios in a variety of life sciences settings:

- **Medical imaging:** A malevolent actor introduces code that produces inconspicuous alterations in medical imaging data that are undetectable by humans yet deceive the diagnostic AI-enabled computerized system into erroneously categorizing severe conditions as benign. This assault jeopardizes patient safety by impairing the attacked system's ability to serve its diagnostic purpose, thereby postponing treatment for life-threatening ailments.

- **Genetic analysis:** An adversarial AI system executes a model inversion attack against another AI used for genetic analysis in customized medicine. The malicious AI reconstructs sensitive genetic data of individuals from the victim AI system's training data set, violating data privacy. This AI-driven breach of privacy may disclose susceptibilities to certain illnesses, contravening data protection laws and jeopardizing patient safety.

- **Pharmaceutical research:** An adversarial assault discreetly contaminates the training data of an AI-enabled computerized system used in drug development. The victim AI's model is compromised, resulting in a preference for certain chemical configurations that are, in fact, less effective or detrimental. This sabotage undermines product quality and patient safety for the further development process, including clinical trials stages, potentially resulting in harm to trial participants.

- **Clinical trials:** An adversarial system executes a membership inference attack on an AI-enabled computerized system that manages clinical trial data. The assailant AI system identifies the individuals involved in a particular study by analyzing the clinical trial data. This breach of confidentiality may expose health information, compromising the integrity of the clinical trial process and contravening ethical standards established by health authorities.

- **Medical devices:** An assailant AI system embeds a backdoor in another AI model used in medical devices such as insulin pumps or pacemakers. The backdoor enables the assailant AI system to induce failures or changes of dosages in the target system managing the medical device.

From a conceptual perspective, the FDA states in their draft guidance [102] on AI-specific risks in the context of medical devices, that various scenarios and cybersecurity threats should be considered:

- Data poisoning: Injecting inauthentic or maliciously modified data
- Model inversion: Inferring details from or replicating models
- Model evasion: Modifying input samples to deceive models
- Data leakage: Leaking sensitive training or inference data
- Overfitting: Exploiting insufficient generalization of models
- Bias: Introducing bias or exploiting known bias
- Performance drift: Altering the underlying data distribution, which degrades model performance

27.8 Cybersecurity Risk Mitigation

Regulated companies should establish a holistic cyber risk control strategy that considers organizational and technical controls as well as human elements. This strategy should raise awareness and include dedicated training.

27.8.1 Cybersecurity Frameworks

Regulated companies may base their cybersecurity strategy by leveraging frameworks such as ISO 27001 [59], ISO/IEC 42001 [5], SOC 1/SOC 2/SOC 3 [66], and HITRUST CSF [130], as outlined in this appendix. They may also consider certifications to codify and assess adherence to such frameworks depending on the use case and context of use.

An orchestrated approach to security posture management can ensure organizational adherence to established internal and external security frameworks, while providing assurance to interested third parties that security supportive processes are established and subject to periodic internal and external audits.

27.8.2 Organizational Aspects

In case of a cybersecurity incident, considerate steps are required, often in a short time. Therefore, clear processes for communication and disclosure should be established, including planning the following aspects of communicating the issue as well as timing for resolution and/or available workarounds:

- **How** (e.g., pre-approved template communication for different scenarios)
- **Why** (assessment by the incident response team)
- **When** (immediate action after recovery, when to notify stakeholders/customer, ongoing updates, and final updates)
- **Who** (clear roles and responsibilities)

Cybersecurity communication should integrate with the business continuity planning to ensure that business operations remain minimally disrupted during a breach or vulnerability management.

Training and providing relevant information to staff and end users is essential for cybersecurity risk mitigation, since cyber attackers may exploit human error (e.g., phishing, social engineering) as a weak link in an organization's security posture. General elements of such training include:

- Phishing awareness
- Social engineering defense
- Password security and authentication practices
- Computer and mobile security
- Data management and confidentiality

Establishing AI literacy is also a means of cybersecurity controls, to allow end users and stakeholders to distinguish usual from unusual behavior of AI-enabled computerized systems. In addition, AI-specific cybersecurity risks should be covered and contextualized in the context of use of AI sub-systems.

Regular cybersecurity updates and refresher training keep employees informed about new threats, emerging trends, and best practices.

Further information on business continuity planning is in Appendix P3.

27.8.3 Use of Tools and Technology

Cybersecurity tools and methods can support an AI-enabled computerized system's cybersecurity posture and in managing cybersecurity risks. In case of involvement of suppliers, general availability of such tools may inform supplier assessment, audits, and contractual items such as SLAs, see Appendix M2.

27.8.3.1 Cybersecurity Management

Compliance management tools that support cross-industry frameworks help achieve and maintain high cybersecurity standards. Third-party tools for actively managing one or multiple third-party security frameworks are widely available.

27.8.3.2 Firewalls and Intrusion Detection and Prevention Systems (IDS/IPS)

AI-enabled computerized systems often rely on cloud resources or distributed computing networks to support large data sets or provide required computing power.

Firewalls and IDS/IPS provide defense for such platforms. An IDS may help in identifying unusual patterns in model training requests or signs of efforts to corrupt the training data or its results.

27.8.3.3 Anti-Virus and Malware

Due to potential corruption of training models and data from viral or malware attacks, AI-specific malware detection systems may support detecting anomalies in model behavior or training data. Data Integrity checks should be implemented to identify instances of potential third-party manipulation.

27.8.3.4 Encrypting Data

Theft of information or data from, or unauthorized access to, AI-enabled computerized systems may result from either poorly encrypted or unencrypted data sets. Trusted execution environments or air- gapped systems can provide additional remediation in addition to the use of encryptions. In some scenarios, homomorphic encryption may be helpful to eliminate the need for decryption of the scoped model. Regulated companies should consider encryption for data both at rest and during transit.

27.8.3.5 Implementation of Access Control and Identity Management Systems

Access controls for all systems involved in the AI-enabled computerized system life cycle are a good practice, while considering further means such as RBACs and multi-factor authentication.

27.8.3.6 Systems for Security Information and Event Management (SIEM)

Some logging and monitoring systems are specifically designed for AI use cases, with particular attention on the inputs, outputs, and performance indicators of the model. Anomaly detection tools can help identify irregular trends in the model's performance or use.

27.8.3.7 Vulnerability Scanning and Penetration Testing Tools

Vulnerability scanning and penetration testing tools can support in the identification of possible loopholes in an automated way, to derive early action.

27.8.3.8 Backup and Recovery

Approaches for Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) should be considered as part of planning for recovery scenarios. Objectives depend on the nature and the complexity of the systems and data, as well as on the sensitivity of outages, e.g., a dynamic system may require different RPO/RTO compared to static system designs.

27.8.3.9 Threat Intelligence

As referenced throughout this appendix, government bodies such as the US Cybersecurity Infrastructure Agency [131] and the Federal Office for Information Security in Germany [95] provide further resources for threat intelligence across the cybersecurity space. These resources, along with the use of third-party applications that provide threat intelligence, can help to achieve an up-to-date view on cybersecurity threats.

28 Appendix S6 – AI in and as Medical Devices

28.1 Introduction

This appendix demonstrates how concepts presented throughout this Guide can be applied to the use of AI in Software in Medical Devices (SiMD) and Software as Medical Devices (SaMD) and their regulatory requirements.

This appendix is aligned with various guidance including Good Machine Learning Practices for Medical Device Development: Guiding principles [115] and the extension regarding Transparency for Machine Learning-enabled Medical Devices [33]. It covers various challenges such as the evaluation of the safety and effectiveness of dynamic system designs or post-market monitoring of AI-enabled devices as outlined by the FDA [132]. Furthermore, it considers guiding principles for Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices [133], and the Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions [134].

This appendix is applicable to medical devices that embed AI, referred to as AI-enabled medical devices. While *ISPE GAMP 5 (Second Edition)* [2] provides guidance on some aspects of medical devices, this appendix addresses the use of AI in medical devices.

This appendix provides guidance to:

- Organizations developing a product whose intended use makes it a medical device
- Medical device manufacturers who intend to use AI as or as part of their medical devices
- Suppliers to medical device manufacturers

28.2 Challenges in the Context of AI-Enabled Medical Devices

When developing AI-enabled medical devices, manufacturers and stakeholders face challenges that can impact their successful deployment and operation. These challenges apply to the full life cycle. In addition to the challenges discussed in Appendix S2, additional aspects specific to AI-enabled medical devices include:

- **Maintaining data privacy:** Organizations need to ensure that Personal Health Information (PHI) and Personally Identifiable Information (PII) are protected throughout the AI-enabled medical device product life cycle.
- **Availability of data sets that are fit for purpose:** Organizations often face data privacy concerns, operating with statutes like privacy and data protection regulations, which may result in barriers in achieving data sets that are fit for purpose for the specific intended use of the AI-enabled medical device.
- **Implementing a proactive risk management strategy:** A proactive risk management strategy aims to support changes and updates to the AI-enabled medical device while it is on the market.
- **Navigating an evolving regulatory landscape:** Organizations face frequent updates and local particularities, which may introduce new requirements or guidelines impacting the AI-enabled medical device life cycle while it is on the market.
- **Ensuring adequate training for the use of the AI-enabled medical device:** Training of end users may fall outside the control of the manufacturer or regulated company.

- **Lack of explainability of models:** This can hinder trust of end users and potentially impact product approval chances.
- **Addressing bias in data and ensuring fairness:** Organizations need to consider outcomes for diverse patient data and populations to prevent disparities in healthcare outcomes.
- **Reliability of AI models:** Model performance may vary based on patient vitals, demographics, and conditions, as well as availability and quality of data under real-world conditions, where organizations do not have full control over the environment where the device is used.
- **Operating in restricted environments:** Managing the demands for data availability, real-time processing, and significant computational power can pose substantial challenges in resource allocation and infrastructure design, particularly for use of AI in distributed devices.
- **Dynamic system designs:** Organizations face challenges when considering the use of dynamic learning, such as changing patient behaviors, even though they may be helpful in adapting to dynamics while on the market.
- **Interoperability:** Organizations may face interoperability challenges when integrating with legacy systems using differing data formats that can disrupt seamless data flow and integration.
- **Legal aspects**: Organizations need to understand the legal implications if a model produces faulty outcomes, which can result in patient harm and lead to issues regarding the organization's and the physicians' liability.

These challenges underscore the need for strategic planning, ongoing evaluation, and adaptation to ensure that AI-enabled medical devices are safe, effective, and compliant throughout the product life cycle.

28.3 General Considerations

Once the organization identifies the potential for their product to be considered a medical device, strategic steps towards market authorization are required. Determination of a product as a medical device depends on its intended use, where performing medical functions such as diagnosing, treating, or preventing diseases leads to a high likelihood of being regulated under the respective market's regulatory jurisdiction.

The following subsections describe key elements of a strategic plan to efficiently advance AI-enabled medical devices to market while ensuring regulatory compliance and reducing the risk of product recalls and penalties.

28.3.1 Understand Regulations and Determine the Device Classification

The organization should familiarize itself with regulatory expectations and guidance for medical devices. The intended use, risk level, and key functionalities will determine the medical device classification.

Regulatory requirements, such as the FDA's medical device classification [135] or classification per Medical Device Regulation [136] for products marketed in the EU, inform the risk and applicable rigor applied throughout the life cycle, while additional requirements can result from using AI in medical devices.

For instance, medical devices of class IIa and above are considered a high-risk system under the EU AI Act [24], with tighter development, documentation, transparency, and further obligations; see EMA reflection paper [43] and its guidance on the use of medical devices in a GCP setting. Design controls requirements (such as 21 CFR Part 820.30 [137]) may also be needed, especially for high-risk medical devices.

Training stakeholders (including those involved in developing, testing, and maintenance, and the Quality Unit) are typically required, in line with regulatory standards, best practices, and the defined QMS processes.

28.3.2 Implementation or Modification of the QMS

When considering AI-enabled medical devices, adopting or modifying existing QMS processes to align with recognized standards such as ISO 13485 [98], IEC 62304 [6], ISO 14971 [7], and ISO/IEC 23894 [8] can facilitate compliance and audit readiness.

Organizations can follow this Guide and its suggested activities to establish or modify processes in their QMS to account for expectations in AI-enabled medical device development and validation, in alignment with applicable regulations. The QMS process should also comply with electronic records and electronic signatures regulations (e.g., 21 CFR Part 11 [89]).

28.3.3 Development and Validation of AI-Enabled Medical Devices

Software life cycle processes should be established to develop and validate AI-enabled medical devices. In this context, AI-enabled medical devices should be designed for:

- Safety: Use risk management framework as suggested by ISO 14971 [7] and ISO/IEC 23894 [8]
- Cybersecurity: Use ISO 27001 [59] and AAMI TIR57 [138], for instance; see Appendix S5
- Usability: Use IEC 62366 [139]

The life cycle activities mentioned in this Guide can be used with IEC 62304 [6] for developing and releasing AI-enabled medical devices. The phases outlined in this Guide, when compared to IEC 62304 are as follows:

- **Concept phase:** Overlaps with IEC 62304 Section 5.1 Software development planning and most of 5.2 Software requirements analysis
- **Project phase:** Overlaps with IEC 62304 Section 5.2 Software requirement analysis, 5.3 Software architectural design, 5.4 Software detailed design, 5.5 Software unit implementation and verification, 5.6 Software integration and integration testing, 5.7 Software system testing, 5.8 Software release, and part of 6.1 Establish software maintenance plan
- **Operation phase:** Overlaps with IEC 62304 Section 6.1 Establish software maintenance plan and 6.2 Problem and modification analysis, while 6.3 Modification implementation is linked to change management activities, which share similarities with project phase activities

28.3.4 Post-Market Surveillance and Maintenance of the AI-Enabled Medical Device

Regulations require ongoing monitoring and maintenance (such as those mentioned in IEC 62304 [6]) to provide software updates, adverse event mitigations, and fixes to cybersecurity vulnerabilities.

The operation phase in this ISPE AI Guide, in conjunction with *ISPE GAMP 5 (Second Edition)* [2], generally covers the software maintenance activities typically required as mentioned in IEC 62304. If there are gaps in processes or activities required for compliance, organizations need to update their QMS.

Post-market surveillance planning may include the collection of data from real-world use to demonstrate ongoing adequacy of the AI-enabled medical device. See Art. 72 of the EU AI Act [24] for an example of such requirements: *"The post-market monitoring system shall actively and systematically collect, document and analyse relevant data which may be provided by deployers or which may be collected through other sources on the performance of high-risk AI systems throughout their lifetime."*

28.3.5 Regulatory Submissions and Interaction with Regulatory Authorities

It is recommended to engage early with regulatory agencies to understand the requirements, for example, during the concept phase (see Appendix P1) to determine the most efficient path to market. Regulated companies may be required to establish formal premarket submissions (especially the ones related to intended use, technical information on the safety and efficacy of the data, and validation results), for example, by means of the FDA's eSTAR program [140]. FDA draft guidance [102] includes an overview of the expected elements of an AI-enabled medical device submission.

A technical file or a device master record documenting risk management, development, and validation activities should be created and maintained for regulatory submissions and for review during audit and regulatory inspections. It is recommended to liaise with regulatory agencies for any submission, audit, or inspection before the product is approved for use on the market.

28.4 Governance

28.4.1 Life Cycle Approach

Organizations are expected to establish a QMS that is compliant with ISO 13485 [98] and in line with risk management activities in ISO 14971 [7], aiming for patient safety and compliance aspects of the AI-enabled medical device throughout its life cycle. See IEC 62304 [6] for guidance on the life cycle design.

A brief description of typical life cycle activities, contextualized for AI-enabled medical devices, is as follows:

- In the **concept phase,** the intended use of the AI-enabled medical device is formally defined, and a rigorous risk assessment performed.

- Focus shifts during the **project phase** to the design and development of the AI-enabled medical device, ensuring it aligns with user needs and regulatory requirements. Verification and validation processes are integral at this stage, serving to confirm that the software meets specifications and intended functions in a safe and effective manner.

- As the AI-enabled medical device transitions into the **operation phase,** ongoing management activities take precedence, including change and configuration management, to maintain the integrity of the software through updates or modifications.

  Deployment is orchestrated to integrate the product into healthcare settings seamlessly, and Post-Market Surveillance (PMS) is employed alongside maintenance practices to perform ongoing monitoring of the software's performance when applied in practice.

- The **retirement phase,** or end-of-life management, governs the device at its end of life, where a structured approach is taken to decommission the software safely, managing data archival and disposal in accordance with legal and ethical standards to protect patient information and maintain data security.

See Section 28.7 for additional information on life cycle phase aspects.

Note: The FDA uses the "total product life cycle" approach [141] to stress the relevance of the medical device life cycle. A dedicated version for generative AI includes the steps Planning and Design, Data Collection and Management, Model Building and Tuning, Verification and Validation, Model Deployment, Operation and Monitoring, Real-World Performance Evaluations, which leads back to Planning and Design. [142]

28.4.2 QRM

Organizations should consider using ISO/IEC 23894 [8], ISO 14971 [7], ISO/TR 24971 [143] and AAMI TIR34971:2023 [92], to support their risk management process and address risks specifically associated with AI-enabled medical devices.

A dedicated or specialized sub-team may augment the risk management team to manage AI-related risks across the medical device's life cycle, as the required expertise may differ from those provided by traditional medical device risk management stakeholders.

Following the "Inclusivity" principle from ISO/IEC 23894 [8], organizations should involve key roles and stakeholders in developing and supporting AI-enabled systems to ensure comprehensive collaboration and participation in risk management activities.

Human behavior and organizational culture can greatly impact all areas of risk management. Therefore, organizations should monitor the cultural landscape to understand how AI-enabled medical devices and their components interact with existing societal patterns. This can reveal potential impacts on equitable outcomes, privacy, fairness, safety, security, environmental factors, and human rights.

Organizations should establish risk management activities specific to AI-enabled medical devices, including the following key elements:

- **Commitment:** Define overarching AI objectives, commit the necessary resources (people, processes, and technology) to risk management activities, and work to continually build stakeholder confidence.
- **Integration:** Integrate risk management activities with core functions and activities involved in developing and deploying AI-enabled medical devices.
- **Design:** Structure the AI risk management process by:
    - Understanding the organizational context, both internal and external
    - Establishing a clear commitment to AI and risk management
    - Assigning AI-related roles with corresponding responsibilities and accountabilities
    - Setting up collaboration and communication methods to ensure that stakeholders are informed and consulted on AI-related risks
- **Implementation:** Implement the risk management process with the necessary controls, ensuring continuous monitoring through internal audits and regular reviews.
- **Evaluation:** Regularly assess the risk management process for adequacy and alignment with objectives.
- **Improvement:** Continually enhance and refine the process to adapt to changing contexts and emerging risks.

28.4.3 Scalable Life Cycle Activities

Organizations should consider the scalability of life cycle activities for AI-enabled medical devices to address varying complexities across devices, while still ensuring processes such as risk management, quality management, and oversight are adapted in line with regulatory expectations.

Furthermore, as device capabilities or served patient populations change, organizations should design life cycle activities to expand or contract in response to the AI-enabled medical device's changing risk profile, complexity, and intended use across markets.

28.4.4 Supplier Management

Medical device QMS standards (such as ISO 13485 [98] and IEC 62304 [6], which explicitly mention requirements on supplier management and Software of Unknown Provenance, respectively) encourage establishing processes to assess conformity with standards and expectations to determine the suitability of all suppliers of services, equipment, and software used in the medical device development.

Medical device organizations are responsible for identifying all suppliers and evaluating the risks their product or service poses to device users and patients. Medical device organizations are also responsible for ensuring that the suppliers comply with the necessary standards and that controls are in place to validate third-party components. Configuration management (e.g. via a Configuration Management Database) and an SBOM support traceability and provide an overview of all software components within the medical device.

Software of Unknown Provenance (SOUP) necessitates rigorous evaluation against defined safety and performance requirements. In the context of AI-enabled medical devices, supplier controls are particularly important when considering foundation models or suppliers that provide ML components to the medical device. Medical device organizations should ensure that supplied components meet unique identifier requirements, if applicable, and that component evaluations are documented in accordance with the company's QMS standards.

Medical device organizations need to ensure suppliers provide thorough records and evidence demonstrating their adherence to quality management standards, such as ISO 13485 [98], IEC 62304, and ISO 14971 [7], and compliance with regulatory requirements. Additional items to consider include communication channels, review of key information items, and involvement in planning activities for further development steps.

28.4.5 Fit for Purpose Data and Data and Model Governance and Management

General aspects related to data and model governance, as described in Appendices M6 and M7, apply. For medical devices, organizations should consider the following aspects in more detail:

- Data collection for ongoing monitoring and improvement of models often requires data use agreements, particularly when dealing with patient data.
- Details on the context of use linked to obtained data may help to identify weaknesses (e.g., low performance of the model) in a targeted manner.
- Model updates need to consider the distributed nature of medical devices, i.e., model repositories and confirmation of model updates need to ensure full traceability of the actual model version in operation in the medical device.
- AI-enabled medical devices may contain feedback loops with end users, including patients and healthcare professionals. Information on the appropriate use, and training where applicable, should be provided to ensure the suitability of information received for ongoing device monitoring and further development.
- If data is provided by external parties, organizations should establish supplier controls including requirements for data quality, provenance, privacy, security, and ethical/legal compliance.
- Use of synthetic data may be helpful to test the robustness of the model and the device and to augment case data sets to expand the coverage of populations or general situations relevant in the context of use; further considerations regarding the suitability of synthetic data approaches are provided in Appendices P2 and M7.

28.4.6 Inspection Readiness and Certifications

In addition to the guidance provided in Appendix M4, medical device development involves further interactions with third parties, e.g., health authorities for audits, inspections, registration, and certifications, and other parties such as notified bodies in the EU.

Design decisions and requirements are of high relevance during the certification process, necessitating rigor in documentation practices. Furthermore, organizations should understand the relevance of validation data for regulatory submissions [120] and specifically 21 CFR Part 820 [137].

Regulatory bodies may examine risk documentation for entire systems that extend beyond the medical device boundaries. It is important to include components that generate, store, transmit, or analyze data within the entire system in the risk assessment, specifically in the hazard analysis and cybersecurity assessment. For instance, while a Laboratory Information Management System (LIMS) might not be defined as part of the device, it plays a critical role in sample tracking and should be assessed for potential patient safety and data integrity risks in the end-to-end system context. In preparation for regulatory inspections, it is advisable to review not only the Design History File (DHF) but also information pertaining to all systems involved in the overall workflow and data flow.

28.5 Design

Medical devices may operate under restricted conditions, e.g., a limited amount of computing power may be available particularly for battery powered devices. Similarly, devices may perform quick decisions in the range of milliseconds, which limits the use of cloud settings. These infrastructure restrictions have implications on aspects of AI-enabled medical devices including the complexity of applicable models and the use of additional features such as XAI methods.

Therefore, the design of the medical device is a crucial step to prepare for efficient implementation as well as safe, robust, and effective behavior in the context of the intended use. Organizations should consider various aspects, including:

- Dedicated roles and responsibilities, with focus on augmented or new roles
- Trustworthy AI
- Dedicated regulatory processes in medical device development

Note: Continual ML (referred to as dynamic systems in this Guide, see key terms in Chapter 2) has potential but presents challenges for medical devices. According to the FDA glossary [120], *"a continual machine learning model has a defined learning process to change its behavior."* The German Notified Bodies Alliance state that dynamic system designs are not considered certifiable *"unless the manufacturer takes measures to ensure the safe operation of the device within the scope of the validation described in the technical documentation."* [144] Therefore, the decision to opt for a dynamic device should be based on robust reasoning with a clearly expected clinical benefit guided by a thorough control strategy on the dynamic learning behavior.

28.5.1 Roles and Responsibilities

In addition to the general roles and responsibilities described in Chapter 6, organizations should consider the following aspects in the context of AI-enabled medical devices:

- When developing a medical device algorithm, effective implementation may require collaboration with equipment/instrument operators, administrators, and third-party software vendors.
- Quality Assurance roles should apply the concepts of ISO 13485 [98], IEC 62304 [6], and ISO 14971 [7] to the unique design, development, testing, and maintenance of an AI-enabled medical device.

- Stakeholders outside the organization, such as clinical staff, healthcare professionals, and patients, should be sufficiently trained or provided with information to allow for effective and safe use of the AI-enabled medical devices.

- Regulatory affairs roles need to be knowledgeable in managing regulatory submissions for the AI-enabled medical device, following current and evolving regulations, guidance, and standards.

- Involvement of stakeholders from engineering teams is important for planning the integration and safety functions related to sensors, hardware, and equipment.

28.5.2 Trustworthy AI

General considerations on trustworthy AI per Appendix M9 apply to AI-enabled medical devices. Further considerations include:

- **Human autonomy and control:** Users should be made aware that they are interacting with AI-enabled medical devices. A stop-function to terminate operation of the AI-enabled medical device is expected. In settings where such decisions may be of high criticality (such as life-threatening situations), safety controls should be embedded to prevent unintended use of such stop functions.

  The possible limited technical background of users is relevant when establishing appropriate safety controls and ensuring effective user interaction.

- **Safety and security:** Organizations should consider the direct impact of medical devices on patients, especially for devices that may be used without the control of healthcare professionals.

  Considering that medical devices can be distributed and operated globally, and the implications of handling patient data, organizations should apply thorough cybersecurity practices; see Section 28.6.

- **Fairness and mitigation of bias:** The coverage of patient populations in the context of the intended use of the medical device is highly relevant.

  In some situations, a careful balance is needed regarding the availability of data, for demonstration of the performance and safety of the device on a representative data set, as well as possible exclusions of patient populations.

  When fully representative data is difficult to procure, using simulated or synthetic data may help reach targets. If simulated or synthetic data is used, the rationale, methodology, stakeholder engagement and approval, and other considerations should be documented.

- **Transparency:** Adequate transparency on the limitations, benefits, and risks of the medical device is expected from organizations. They may also be required to provide transparency on their development approach and rationale for the use of AI, commensurate with the expected AI literacy of patients or healthcare professionals.

- **Accountability:** The use of medical devices can vary; sometimes, trained healthcare professionals make the final decisions, while other AI-enabled medical devices are used with some degree of autonomy by patients.

  If users are required to interpret the model's output, it is essential that usability factors and risks be defined, evaluated, and integrated into the device design. Given the specific context of use, a clear understanding of accountability and responsibilities to execute effective human oversight should be established (e.g., in the instructions for use).

- **Privacy and data protection:** Organizations need to consider the highly sensitive nature of patient data managed in AI-enabled medical devices. This includes the use and handling of data generated via feedback loops from patients or healthcare professionals. Organizations should consider agreements regarding the use of data for further development purposes.

- **Sustainability:** While sustainability considerations in the context of medical devices are typically more concerned with hardware components, good practices also involve the potential reuse of models and data, in alignment with applicable data privacy regulations.

28.5.3 Quality Unit Oversight

The Quality Unit plays a dedicated role in ensuring that expectations for the safety and effectiveness of medical devices are met. Relevant aspects of AI-enabled medical devices for the Quality Unit include:

- Involvement in design reviews to support the holistic coverage of risks and suitability of risk mitigation and control measures, targeting effective and safe devices.

- Oversight of clinical trials to support adherence of quality standards and regulatory expectations.

- Review of DHF documentation specific to AI-enabled medical systems. Keeping the need for transparency in mind, the Quality Unit should be engaged from the early stages of the project through maintenance advising the project team on documentation needs, which may vary from project to project.

- Quality assurance of documentation (e.g., DHF) and submissions (e.g., eSTAR) relevant for certification processes and regulatory submissions.

The Quality Unit requires appropriate AI literacy to execute their oversight activities responsibly and effectively.

28.6 Cybersecurity Aspects of AI in Medical Devices

Organizations should consider that increased connectivity has integrated individual medical devices into larger systems, including healthcare networks and servers as well as cloud environments. This interconnectedness means that cybersecurity threats to any component in the system can impact the safety and effectiveness of a device. Therefore, organizations should ensure device security not only from the point of view of the medical device, but also considering its role within the broader system, including platforms and connected systems or software products and third-party software components.

As cybersecurity is part of device safety and effectiveness, organizations should also implement cybersecurity controls during premarket development, taking into account the intended and actual use environment. Device cybersecurity design and documentation are expected to scale with the cybersecurity risk of that device.

The FDA requires key elements to be created and submitted as part of the medical device's DHF, as discussed in the guidance on Cybersecurity in Medical Devices [127]. As stated in this FDA guidance, *"Section 524B(c) of the FD&C Act defines 'cyber device' as a device that (1) includes software validated, installed, or authorized by the sponsor as a device or in a device; (2) has the ability to connect to the internet; and (3) contains any such technological characteristics validated, installed, or authorized by the sponsor that could be vulnerable to cybersecurity threats."* While cyber devices generally exhibit possible exposure to cyber threats, AI-enabled devices that are simultaneously cyber devices exhibit further attach vectors to cyber threats; see Appendix S5.

Section 524B(c) of the FD&C Act defines “cyber device” as a device that organizations should define post-market processes and activities in a Cyber Security Management plan to ensure ongoing cyber security risk management and include the following:

- Personnel responsible
- Sources, methods, and frequency for monitoring and identifying vulnerabilities
- Periodic security testing
- Timeline to develop and release patches and update processes
- Coordinated vulnerability disclosure process and communication of forthcoming remediations, patches, and updates to customers and, as necessary, regulatory bodies.

Additionally, for Premarket Approval (PMA) devices with periodic reporting requirements under 21 CFR 814.84 [145], information concerning cybersecurity vulnerabilities, as well as device changes and compensating controls implemented in response to this information should be reported to the FDA in a periodic annual report.

Besides references provided in Appendix S7, the following documents may be relevant regarding the cyber security aspects of AI-enabled devices, at the time of writing:

- FDA Guidance for Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions [127]
- FDA Final Guidance: Post-Market Management of Cybersecurity in Medical Devices [128]
- FDA Guidance: Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software [146]
- FDA Draft Guidance: Select Updates for the Premarket Cybersecurity Guidance: Section 524B of the FD&C Act [147]
- NIST SP 800-30 Rev 1 Guide for Conducting Risk Assessments [147]
- MDCG 2019-16 Guidance on Cybersecurity for medical devices [149]
- IMDRF/CYBER WG/N60:2020 Principles and Practices for Medical Device Cybersecurity [150]
- TEAM NB position paper on Cybersecurity [151]
- IG-NB (German Notified Bodies Alliance) – Questionnaire “Cybersecurity for Medical Devices – Technical Documentation” [152]
- IG-NB (German Notified Bodies Alliance) – Questionnaire “Cybersecurity for Medical Devices – Audit” [153]
- BS EN IEC 81001-5-1:2022 Health software and health IT systems safety, effectiveness and security [154]

28.7 Life Cycle Guidance

This section provides further details and contextualization of general guidance in life cycle Appendices P1, P2 (design and development, and verification and validation), P3 (maintenance, change and configuration management, and post-market surveillance), and P4 (retirement and decommissioning).

28.7.1 Concept Phase

Organizations should consider the following aspects when applying guidance in the concept phase for AI-enabled medical devices:

- The business need or opportunity should be considered from a patient perspective, including unmet clinical needs.
- Dedicated regulations and certification requirements, as well as the market authorization strategy, need to be considered when planning subsequent activities in developing the AI-enabled medical device.
- The context of use includes stakeholders that may hold a low AI literacy; this informs the risk management activities as well as requirements for the AI-enabled medical device.
- Suppliers should be knowledgeable about the particularities of medical device development, including standards such as ISO 13485 [98].

28.7.2 Project Phase

Organizations should consider the following aspects when applying guidance in the project phase for AI-enabled medical devices:

- A Design Development Plan should be developed to outline the design inputs, focusing on regulatory compliance and user needs.
- The interplay of hardware and models, when facing restrictions in computing power, needs to be considered for the development of the device. Prototyping is commonly applied to test the effectiveness of the AI-enabled medical device design, that typically goes beyond prototyping to assess feasibility as suggested in the concept phase (Appendix P1).
- Scenarios of real-world use may be less controlled than in other GxP areas. The robustness and usability of the AI-enabled medical device to various scenarios needs to be considered in the system design, commensurate with the limited possibilities for training.
- Clinical activities may require demonstration of fitness for intended use to achieve market authorization.[25]
- Regulatory bodies typically expect detailed documentation of project phase activities as part of formal submissions, including product specifications, configuration management, and testing.
- A Predetermined Change Control Plan (PCCP) (e.g., FDA guidance [134]) may be considered, which lists the planned changes to the device once approved and in operation.

[25] In addition to verification of specifications, medical devices may need to undergo clinical validation before receiving market authorization, pertaining to the respective regulatory jurisdiction and the classification of the device. These activities include the determination of whether the product meets requirements and specifications, and the creation of evidence to use for assessing an appropriate balance of benefits and risks, in comparison to alternative therapeutic options available. Data generated during clinical validation yields a (clinical) validation set for this purpose. This data set should not be confused with the (data science related) validation data set, used to evaluate a trained or engineered model during the iterative development process. In this context, the test data set most closely relates to the concept of verification, as designed to achieve an uninformed estimate of model performance compared to expectations and requirements.

28.7.3 Operation Phase

Organizations should consider the following aspects when applying guidance in the operation phase for AI-enabled medical devices:

- Organizations need to implement post-market surveillance processes to ensure the ongoing safety and effectiveness of the AI-enabled medical device, including collection and analysis of real-world data and end-user feedback as well as monitoring of public registers.
- Post-market clinical follow-up activities may be required.
- Formal change management processes may require re-certification of the AI-enabled medical device and notifications or submissions to regulatory agencies and other bodies, e.g. as per 510(k) submission (see for example “Deciding When to Submit a 510(k) for a Software Change to an Existing Device” [156]).

28.7.4 Retirement Phase

If an AI-enabled medical device is being retired (i.e. decommissioned, shut down in a controlled manner, or fully disposed of), a clear, documented retirement plan should be developed, executed, and its execution verified. In addition to the details provided in Appendix P4 about planning and execution of retirement activities, the following key considerations apply when a medical device is being retired:

- **Stakeholder communication and training**
  - Alignment and communication to relevant stakeholders, including healthcare staff and patients, about the retirement activities, timeline, and any required actions
  - Planning and conduct of any training required to incorporate the transition to the superseding medical device or introduce major changes to the system. The training should also cover possible issues or queries that end users and patients might encounter during the conduct of retirement activities.
- **Regulatory activities**
  - Notifying regulatory bodies and following the necessary regulatory activities including delisting the devices
  - Retirement involves disposing of hardware components, where applicable environmental guidelines and requirements should be considered.
- **Record retention**
  - Development and execution of a record retention plan as per medical device record retention requirements
  - For data that does not need retention, data should be deleted in a secure way to protect privacy and prevent unauthorized access, as applicable. A balance should be met regarding further use of data, in compliance with relevant data privacy regulations.

29 Appendix S7 – Statutes, Regulations, Standards, and Guidance

The statutes, regulations, guidance, and regulatory initiatives with dedicated AI relevance presented in Table 29.1[26] were considered during the development of this Guide. Additional general guidance and regulations are also applicable in the broader scope of computerized systems.

Initiatives are sorted by date, descending, as of the time of writing.

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives

Agency	Type	Publication	Name of the Publication	Comment
FDA	Draft Guidance	January 2025	Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products [55]	The draft guidance document introduces a seven-step process, named a risk-based credibility assessment framework. It includes considerations on the question of interest, the context of use of an AI model, AI model risks, a plan to establish AI model credibility, execution of the plan, documentation, and determining adequacy.
FDA	Guidance Document	December 2024	Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions [83]	This guidance document covers PCCPs for AI-enabled (medical) device software functions, i.e., necessary policies, and activities to allow for changes in such devices without the need of a resubmission.
FDA	Executive Summary	November 2024	Total Product Lifecycle Considerations for Generative AI-Enabled Devices [142]	Considerations on the use of Generative AI in medical devices, including regulatory oversight, challenges, application of a risk-based approach, and generating scientific evidence.
EMA	Reflection Paper	September 2024	Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle [43]	Overview of use case areas in the medicinal product life cycle, and key considerations regarding regulatory interactions, technical aspects, governance, integrity, data protection, as well as ethical aspects and trustworthy AI.

[26] Some comments are based on the *ISPE eClinical Good Practice Guide (Second Edition)* [58].

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives (continued)

Agency	Type	Publication	Name of the Publication	Comment
FDA	Glossary	September 2024	FDA Digital Health and Artificial Intelligence Glossary – Educational Resource [120]	Glossary providing definitions, and links to further sources, on key terms with primary focus on AI in medical devices.
NIST	Framework	July 2024	NIST Trustworthy and Responsible AI NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile [93]	Cross-sectoral profile as a companion to the AI Risk Management Framework [112], i.e., an implementation of Risk Management Framework Functions with a dedicated focus on Generative AI.
TGA	Guidance	June 2024	National framework for the assurance of artificial intelligence in government [157]	Foundations for a nationally consistent approach to AI assurance, providing expectations and aiming for consistency and certainty for partner relationships; assistance to develop, procure, and deploy AI in a safe and responsible way.
European Parliament	Regulation	June 2024	Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence [24]	Horizontal regulation that poses requirements on the use of AI, based on four levels of risk: unacceptable, high risk, limited risk, minimal risk; it has direct implications on medical devices which may be considered as high-risk cases, and other GxP areas.
FDA / Health Canada / MHRA	Guidance	June 2024	Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles [33]	Extension to the Good Machine Learning Practices [115] that focuses on the role of transparency, including guidance to ensure that stakeholders receive information they need in the context of AI-enabled medical devices.
White House	Executive Order[27]	October 2023	Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence [46]	Order acknowledging the potential of use of AI across sectors, and the urgency of governing the development and use of AI in a safe and responsible manner.

[27] The order was revoked in 2025 and is listed here for historical purposes, providing an overview of various approaches to define AI and AI-enabled systems.

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives (continued)

Agency	Type	Publication	Name of the Publication	Comment
TGA	Regulation	August 2023	Regulation of Software Based Medical Devices [158]	Guidance seeks to help manufacturers and sponsors understand how the TGA interprets requirements, and thus indicates how manufacturers and sponsors can comply; it includes considerations on AI text-based products like LLMs.
EU	Regulation	June 2023	Machinery Regulation (EU) 2023/1230 [118]	Harmonization of health and safety requirements for machinery in all Member States and to remove obstacles to trade in machinery between Member States; it mentions AI briefly as part of advanced machinery approaches.
MHRA	Guidance	June 2023	Software and Artificial Intelligence as a Medical Device [159]	The roadmap mentions three key areas to promote responsible innovation for medical device software: requirements to provide assurance for acceptably safe and working as intended devices, clarity of requirements including guidance and processes, and considerations of further stakeholders in the medical device ecosystem such as National Institute for Health and Care Excellence and NHS England.
FDA	Discussion Paper	May 2023	Artificial Intelligence in Drug Manufacturing [44]	The discussion paper elaborates on the potential of AI to improve drug manufacturing efficiency, quality control, while maintaining regulatory compliance; it encouraged industry feedback on best practices and needs for regulatory clarification.
FDA	Discussion Paper	May 2023	Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products [155]	The discussion paper provides a) a landscape of current (at the time of writing) and potential use cases of AI and ML, b) considerations on the use of AI, and c) an outlook on next steps and stakeholder engagement; specific questions were designed to elicit feedback and focus points from industry.

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives (continued)

Agency	Type	Publication	Name of the Publication	Comment
EMA	Guidance	March 2023	Guideline on Computerised Systems and Electronic Data in Clinical Trials [108]	Guidance on the use of computerized systems in clinical trials; requirements regarding AI beyond generally applicable expectations to all systems are not covered in this ISPE AI Guide in this version; however, the Guide acknowledges its increasing importance in clinical trials.
Department for Science, Innovation & Technology (UK)	Policy Paper	March 2023	A pro-innovative approach to AI regulation [161]	Announcement of developing a framework to bring clarity and coherence to the AI regulatory landscape, underpinned by key principles: • Safety, security, and robustness • Appropriate transparency and explainability • Fairness • Accountability and governance • Contestability and redress
FDA	Guidance	January 2023	Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products [162]	While the guidance primarily focuses on externally controlled clinical trials agnostic of the use of AI, it provides valuable considerations on potential bias, specifically when considering real-world data.
NIST	Guidance	January 2023	AI Risk Management Framework (AI RMF 1.0) [112]	The framework provides guidance on identifying, evaluating, and managing risks for AI systems and their safety, compliance, and effective risk mitigation practices.
ICH	Guidance	January 2023	Quality Risk Management Q9(R1) [31]	The guideline provides general guidance on the principles and practices for QRM in pharmaceutical manufacturing.
FDA	US Law	December 2022	Food and Drug Omnibus Reform Act (FDORA) 2022 [163]	FDORA requires FDA to develop several different kinds of informational documents, including public reports, reports to Congress, strategic plans, and others, including cybersecurity of medical devices.
EMA	Guidance	November 2022	Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data sources [164]	The guidance provides recommendations for the use of the EU metadata catalogue to identify real-world data sources suitable for research questions.

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives (continued)

Agency	Type	Publication	Name of the Publication	Comment
FDA	Presentation	October 2022	Artificial Intelligence/ Machine Learning (AI/ML)-Enabled Medical Devices: Tailoring a Regulatory Framework to Encourage Responsible Innovation in AI/ML [165]	The overview presentation covers AI and ML in medical devices, including use cases, opportunities, challenges, and further information on a proposed regulatory framework.
EMA	Concept Paper	September 2022	Concept Paper on the revision of Annex 11 of the guidelines on Good Manufacturing Practice for medicinal products – Computerized Systems [166]	The concept paper provides information on the drivers of EU GMP Annex 11's revision, listing ML as one of the focus areas to provide further regulatory clarification.
FDA / Health Canada / MHRA	Guidance	October 2021	Good Machine Learning Practice for Medical Device Development: Guiding Principles [115]	The document provides ten general guiding principles for Good Machine Learning Practices for medical device development.
FDA	Action Plan	January 2021	Artificial Intelligence/ Machine Learning (AI/ ML) – Based Software as a Medical Device (SaMD) Action Plan [167]	The document describes a five-point action plan to summarize feedback received from stakeholders regarding the FDA's AI/ML Software as Medical Device Proposed Framework.
White House	Executive Order	December 2020	Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government [168]	The executive order reflects on the role of AI to drive growth and improve quality of life in the US; key areas include Policy, Principles of Use of AI in Government, Implementation of Principles, Agency Inventory of AI Use Cases, Interagency Coordination, AI Implementation Expertise, Responsible Agencies and Officials, Scope of Application, and General Provisions.
European Commission	White Paper	February 2020	White Paper on Artificial Intelligence – A European approach to excellence and trust [169]	Considerations on AI in the political guidelines as a European Coordinated Approach on the human and ethical implications of AI; expresses the commitment to enabling scientific breakthrough and ensuring that new technologies improve citizens' lives while respecting their rights.

Table 29.1: Overview of Statutes, Regulations, Guidance, and Regulatory Initiatives (continued)

Agency	Type	Publication	Name of the Publication	Comment
FDA	Proposal	April 2019	Proposed Regulatory Framework for Modifications to Artificial Intelligence/ Machine Learning (AI/ ML) – Based Software as a Medical Device (SaMD) [170]	Proposal includes considerations on a framework of pre-approved algorithm changes by submission, to acknowledge the possibilities of models and algorithms in medical devices to learn from new data.
European Commission	Guidance	April 2019	Ethics Guidelines for Trustworthy AI [28]	Set of principles that should generally guide the use of AI in ensuring trustworthy AI as lawful, ethical, and robust.
European Commission	Communication	April 2019	Building Trust in Human-Centric Artificial Intelligence [171]	The communication provides a summary of various activities already executed (e.g., development of guidelines for trustworthy AI drafted by the high-level expert group and its plan to achieve international AI ethics guidelines).
FDA	Guidance	December 2018	Real-World Evidence Program [172]	Framework work evaluating the potential use of real-world evidence to support the approval of new drugs.

30 Appendix G1 – References

1. *ISPE GAMP® RDI Good Practice Guide: Data Integrity by Design,* International Society for Pharmaceutical Engineering (ISPE), First Edition, October 2020, www.ispe.org.

2. *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems,* International Society for Pharmaceutical Engineering (ISPE), Second Edition, July 2022, www.ispe.org.

3. *ISPE GAMP® Guide Series,* International Society for Pharmaceutical Engineering (ISPE), www.ispe.org.

4. ISO 9001:2015 Quality Management Systems — Requirements, International Organization for Standardization (ISO), www.iso.org.

5. ISO/IEC 42001:2023 Information technology – Artificial Intelligence – Management System, International Organization for Standardization (ISO), www.iso.org, and International Electrotechnical Commission (IEC), www.iec.ch.

6. IEC 62304:2006 Medical device software — Software life cycle processes, International Electrotechnical Commission (IEC), www.iec.ch.

7. ISO 14971:2019 Medical devices — Application of risk management to medical devices, International Organization for Standardization (ISO), www.iso.org.

8. ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management, International Organization for Standardization (ISO), www.iso.org, and International Electrotechnical Commission (IEC), www.iec.ch.

9. ISO/IEC/IEEE 12207:2017 Systems and software engineering — Software life cycle processes, International Organization for Standardization (ISO), www.iso.org, International Electrotechnical Commission (IEC), www.iec.ch, Institute of Electrical and Electronics (IEEE), www.ieee.org.

10. ISACA® Capability Maturity Model Integration® (CMMI), https://cmmiinstitute.com/.

11. Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., Plöd, M., "CRISP-ML(Q). The Life Cycle Process," Accessed 25 June 2025, https://ml-ops.org/content/crisp-ml.

12. ITIL® Foundation, ITIL 4 Edition, London, UK: Axelos, 2019, www.axelos.com.

13. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

14. International Organization for Standardization (ISO), www.ISO.org.

15. *ISPE GAMP® Guide: Records and Data Integrity,* International Society for Pharmaceutical Engineering (ISPE), First Edition, March 2017, www.ispe.org.

16. ASTM Standard E2500-13 Standard Guide for Specification, Design, and Verification of Pharmaceutical and Biopharmaceutical Manufacturing Systems and Equipment, ASTM International, West Conshohocken, PA, www.astm.org.

17. Federal Food, Drug, and Cosmetic Act (FD&C Act), US Food and Drug Administration (FDA), www.fda.gov.

18. US Public Health Service Act (PHSA), Title 42 of the United States Code (The Public Health and Welfare), Chapter 6A (Public Health Service), www.ecfr.gov.

19. US Food and Drug Administration (FDA), www.fda.gov.

20. European Union (EU), https://european-union.europa.eu/index_en.

21. Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency.

22. Ministry of Health, Labour and Welfare (MHLW), www.mhlw.go.jp/english/index.html.

23. Prescription Drug Marketing Act of 1987; Prescription Drug Amendments of 1992; Policies, Requirements, and Administrative Procedures, A Rule by the Health and Human Services Department, and the Food and Drug Administration, Publication date: 3 December 1999, Effective date: 4 December 2000, www.ecfr.gov.

24. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 12 July 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

25. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 4 May 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj.

26. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, US Department of Health and Human Services, www.hhs.gov/hipaa/index.html.

27. SEC Sarbanes-Oxley Act of 2002, US Securities and Exchange Commission (SEC), www.sec.gov/about/laws/soa2002.pdf.

28. Ethics guidelines for trustworthy AI, European Commission, 8 April 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

29. Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts, European Commission, 21 April 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

30. Vidstrup A., "New EU AI Regulation and GAMP 5," *Pharmaceutical Engineering,* September/October 2023, Vol. 43, No. 5, pp. 56–60, www.ispe.org.

31. ICH Harmonised Guideline: *Quality Risk Management – Q9(R1),* Step 4, 18 January 2023, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

32. Blumenthal, R., Erdmann, N., Heitmann, M., Lemettinen, A.-L., Stockton, B.M., "Machine Learning Risk and Control Framework," *Pharmaceutical Engineering,* January/February 2024, Vol. 44, No. 1, pp. 12–22, www.ispe.org.

33. FDA/Health Canada/MHRA Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles, June 2024, US Food and Drug Administration (FDA), www.fda.gov, Health Canada, www.canada.ca/en/health-canada.html, United Kingdom Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency.

34. Altrabsheh, E., Heitmann, M., Lochbronner, A., "AI Governance and QA Framework: AI Governance Process Design," *Pharmaceutical Engineering,* July/August 2022, Vol. 42, No. 4, pp. 47–53, www.ispe.org.

35. ICH Harmonised Guideline: *Pharmaceutical Quality System – Q10,* June 2008, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

36. ICH Harmonised Draft Guideline: *Guideline for Good Clinical Practice E6(R3, Annex 2),* November 2024, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

37. ICH Harmonised Guideline: *General Considerations for Clinical Studies ICH E8(R1),* Final Version, October 2021, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

38. *ISPE Good Practice Guide: Knowledge Management in the Pharmaceutical Industry,* International Society for Pharmaceutical Engineering (ISPE), First Edition, May 2021, www.ispe.org.

39. American Productivity & Quality Center (APQC), www.apqc.org.

40. Kelley, K., "What is Data Science? A Comprehensive Guide", Caltech Bootcamp, 12 July 2023, https://pg-p.ctme.caltech.edu/blog/data-science/what-is-data-science.

41. Copeland, B. J., "Definition of artificial intelligence," Britannica, 17 June 2025, www.britannica.com/technology/artificial-intelligence.

42. Grobelnik, M., Perset, K., Russell, S., "What is AI? Can you make a clear distinction between AI and non-AI systems?" 6 March 2024, https://oecd.ai/en/wonk/definition.

43. EMA Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle, European Medicines Agency (EMA), 9 September 2024, www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf.

44. FDA Discussion Paper: Artificial Intelligence in Drug Manufacturing, US Food and Drug Administration (FDA), 2023, www.fda.gov/media/165743/download?attachment.

45. "What is artificial intelligence and how is it used?" European Parliament Topics, 20 June 2023, www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used.

46. Federal Register (US), Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 30 October 2023, Revoked January 20, 2025, www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

47. Annex to the Communication to the Commission: Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6 February 2025, https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

48. "Machine learning (ML): All there is to know," International Organization for Standardization (ISO), Accessed 17 June 2025, www.iso.org/artificial-intelligence/machine-learning.

49. "What is machine learning (ML)?" IBM Topics, 22 September 2021, www.ibm.com/topics/machine-learning.

50. PIC/S Guidance: PI 011-3 Good Practices for Computerised Systems in Regulated "GXP" Environments, Pharmaceutical Inspection Co-operation Scheme (PIC/S), 25 September 2007, www.picscheme.org.

51. "Manifesto for Agile Software Development," 2001, https://agilemanifesto.org/.

52. Glossary of Computer System Software Development Terminology (8/95), US Food and Drug Administration (FDA), 11 June 2014, www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/glossary-computer-system-software-development-terminology-895.

53. Medical Devices; Validated Instructions for Use and Validation Data Requirements for Certain Reusable Medical Devices in Premarket Notifications, US Food and Drug Administration (FDA), 9 June 2017, www.federalregister.gov/documents/2017/06/09/2017-12007/medical-devices-validated-instructions-for-use-and-validation-data-requirements-for-certain-reusable.

54. "What is data science?" IBM Topics, 21 September 2021, www.ibm.com/topics/data-science.

55. FDA Draft Guidance for Industry and Other Interested Parties: Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products, US Food and Drug Administration (FDA), January 2025, www.fda.gov/media/184830/download.

56. "Australia's AI Ethics Principles," Department of Industry, Science and Resources, Australian Government, Accessed 17 June 2025, www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles.

57. *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems,* International Society for Pharmaceutical Engineering (ISPE), First Edition, January 2010, www.ispe.org.

58. *ISPE GAMP® Good Practice Guide: Validation and Compliance of Computerized GCP Systems and Data – Good eClinical Practice,* International Society for Pharmaceutical Engineering (ISPE), Second Edition, July 2024, www.ispe.org.

59. ISO/IEC 27001:2022 Information Security, Cybersecurity And Privacy Protection - Information Security Management Systems – Requirements, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

60. ISO/IEC 27002:2022 Information security, cybersecurity and privacy protection — Information security controls, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

61. ISO/IEC 27017:2015 Information technology — Security techniques — Code of practice for information security controls based on ISO/IEC 27002 for cloud services, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

62. ISO/IEC 27018:2019 Information technology — Security techniques — Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

63. ISO/IEC 27701:2019 Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

64. ISO 27799:2016 Health informatics — Information security management in health using ISO/IEC 27002, International Organization for Standardization (ISO), www.iso.org.

65. "Understanding IEC 62443," Editorial Team, International Electrotechnical Commission (IEC), 26 February 2021, www.iec.ch/blog/understanding-iec-62443.

66. Association of International Certified Professional Accountants (AICPA), www.aicpa-cima.com/home.

67. The NIST Cybersecurity Framework (CSF) 2.0, National Institutes of Standards and Technology (NIST), 26 February 2024, doi.org/10.6028/NIST.CSWP.29, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf.

68. ISO/IEC/IEEE 29148:2018 – Systems and software engineering — Life cycle processes — Requirements engineering, International Organization for Standardization (ISO), www.iso.org, and International Electrotechnical Commission (IEC), www.iec.ch, and Institute of Electrical and Electronics (IEEE), www.ieee.org.

69. Habibullah, K., Gay, G., and Horkoff, J., "Non-functional requirements for machine learning: understanding current use and challenges among practitioners," *Requirements Engineering,* 7 January 2023, Volume 28, pp. 283–316, https://link.springer.com/article/10.1007/s00766-022-00395-3.

70. "Classification: Accuracy, recall, precision, and related metrics," Google for Developers, Accessed 18 June 2025, https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall.

71. CS273: Algorithms for Structure and Motion in Biology, Handout #8, 22 April 2003, Stanford University, https://web.stanford.edu/class/cs273/scribing/2004/class8/scribe8.pdf.

72. OECD.AI Catalogue of Tools & Metrics for Trustworthy AI, Organisation for Economic Co-operation and Development (OECD), https://oecd.ai/en/catalogue/metrics/dice-score.

73. Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J., "BLEU: a Method for Automatic Evaluation of Machine Translation," Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311–328, https://aclanthology.org/P02-1040.pdf.

74. Lin, C.-Y., "ROUGE: A Package for Automatic Evaluation of Summaries," University of Southern California, https://aclanthology.org/W04-1013.pdf.

75. Kremers, W.K., "Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time," Technical Report Series #80, Mayo Clinic, April 2007, www.mayo.edu/research/documents/biostat-80pdf/doc-10027891.

76. James, G., Witten, D., Hastie, T., and Tibshirani, R., *An Introduction to Statistical Learning with Applications in R,* Second Edition, NY, NY: Springer, 2021, https://link.springer.com/book/10.1007/978-1-0716-1418-1#overview.

77. ISO/IEC/IEEE 90003:2018 - Software engineering — Guidelines for the application of ISO 9001:2015 to computer software, International Organization for Standardization (ISO), www.iso.org, International Electrotechnical Commission (IEC), www.iec.ch, and Institute of Electrical and Electronics (IEEE), www.ieee.org.

78. IEEE Standards Association, Institute of Electrical and Electronics (IEEE), www.ieee.org.

79. ISTQB® (International Software Testing Qualifications Board), www.istqb.org.

80. NIST AI Test, Evaluation, Validation and Verification (TEVV), National Institutes of Standards and Technology (NIST), Accessed 21 June 2025, www.nist.gov/ai-test-evaluation-validation-and-verification-tevv.

81. "Defining AI incidents and related terms," OECD Artificial Intelligence Papers No. 16, Organisation for Economic Co-operation and Development (OECD), May 2024, www.oecd.org/en/publications/defining-ai-incidents-and-related-terms_d1a8d965-en.html.

82. ICH guideline Q10 on pharmaceutical quality system, European Medical Agency (EMA), September 2015, www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-guideline-q10-pharmaceutical-quality-system-step-5_en.pdf.

83. FDA Guidance for Industry and Food and Drug Administration Staff: Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions, US Food and Drug Administration (FDA), 4 December 2024, www.fda.gov/media/166704/download.

84. ICH Harmonised Guideline: Pharmaceutical Development – Q8(R2), Step 5, August 2009, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), www.ich.org.

85. *ISPE GAMP® Good Practice Guide: Enabling Innovation – Critical Thinking, Agile, IT Service Management,* International Society for Pharmaceutical Engineering (ISPE), First Edition, September 2021, www.ispe.org.

86. ISO 19011:2018 Guidelines for auditing management systems, International Organization for Standardization (ISO), www.iso.org.

87. Perez, A.D., Canterbury, J., Hansen, E., Samardelis, J.S., Longden, H., Rambo, R.L., "Application of the SOC 2+ Process to Assessment of GxP Suppliers of IT Services," *Pharmaceutical Engineering,* July/August 2019, Vol. 39, No. 4, pp. 14–20, www.ispe.org.

88. *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Testing of GxP Systems,* International Society for Pharmaceutical Engineering (ISPE), Second Edition, December 2012, www.ispe.org.

89. 21 CFR Part 11 – Electronic Records; Electronic Signatures, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

90. EudraLex Volume 4 – Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en.

91. ISO 31000:2018 Risk management — Guidelines, International Organization for Standardization (ISO), www.iso.org.

92. AAMI TIR34971:2023 Application of ISO 14971 to machine learning in artificial intelligence—Guide, Association for the Advancement of Medical Instrumentation (AAMI), www.aami.org.

93. NIST Trustworthy and Responsible AI NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, National Institutes of Standards and Technology (NIST), July 2024, doi.org/10.6028/NIST.AI.600-1, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

94. Generative AI Models – Opportunities and Risks for Industry and Authorities, Federal Office of Information Security (BSI), Version 2.0, January 2025, www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.html.

95. Federal Office of Information Security (BSI, Germany), www.bsi.bund.de/EN/Home/home_node.html.

96. Steimers, A., Schneider M., "Sources of Risk of AI Systems," *International Journal of Environmental Research and Public Health,* 18 March 2022, 19(6):3641, www.ncbi.nlm.nih.gov/pmc/articles/PMC8951316.

97. Kaestner, C., "The illustrated guide to software as a medical device (SaMD), IEC 82304-1 and AI," Medical Device HQ®, 8 February 2024, http://medicaldevicehq.com/articles/the-illustrated-guide-to-software-as-a-medical-device-samd-iec-82304-1-and-ai.

98. ISO 13485:2016 Medical devices — Quality management systems — Requirements for regulatory purposes, International Organization for Standardization (ISO), www.iso.org.

99. *ISPE Baseline® Pharmaceutical Engineering Guide, Volume 8 – Pharma 4.0™,* International Society for Pharmaceutical Engineering (ISPE), First Edition, December 2023, www.ispe.org.

100. *ISPE Guide Series: Advancing Pharmaceutical Quality, Cultural Excellence,* International Society for Pharmaceutical Engineering (ISPE), First Edition, November 2022, www.ispe.org.

101. OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications, Organisation for Economic Co-operation and Development (OECD), April 2004, https://unesdoc.unesco.org/ark:/48223/pf0000136561.

102. FDA Draft Guidance for Industry and Food and Drug Administration Staff: Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations, US Food and Drug Administration (FDA), 7 January 2025, www.fda.gov/media/184856/download.

103. ISO/IEC 25012:2008 Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

104. Wang, R.Y., Strong, D. M., “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems,* Spring 1996, Vol. 12, No. 4, pp. 5–23, www.jstor.org/stable/40398176.

105. Myers, D., Blake, B.P., “An Evaluation of the Conformed Dimensions of Data Quality in Application to an Existing Information Quality Privacy-Trust Research Framework (Research-in-Progress),” MIT International Conference on Information Quality, UA Little Rock, October 2017, https://dqmatters.com/_download/P26-ICIQ2017-ConformedDimenstionsofDQ.pdf.

106. Black, A., van Nederpelt, P., “Dictionary of dimensions of data quality (3DQ),” DAMA FL Foundation, 15 November 2020, www.dama-nl.org/wp-content/uploads/2020/11/3DQ-Dictionary-of-Dimensions-of-Data-Quality-version-1.2-d.d.-14-Nov-2020.pdf.

107. FDA Draft Guidance for Industry: Data Integrity and Compliance With CGMP, US Food and Drug Administration (FDA), April 2016, www.fda.gov/files/drugs/published/Data-Integrity-and-Compliance-With-Current-Good-Manufacturing-Practice-Guidance-for-Industry.pdf.

108. EMA Guideline on computerised systems and electronic data in clinical trials, European Medical Agency (EMA), 9 March 2023, www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf.

109. Wyn, S., Perez, A. D., Reid, C., Watson, H., “A GAMP® Approach to Computerized System Life Cycle and IT Process Records,” *Pharmaceutical Engineering,* January/February 2025, Vol. 45, No. 1, pp. 21–25, www.ispe.org.

110. MHRA ‘GXP’ Data Integrity Guidance and Definitions, Medicines & Healthcare products Regulatory Agency (MHRA), March 2018, www.gov.uk/government/publications/guidance-on-gxp-data-integrity.

111. *ISPE GAMP® Good Practice Guide: IT Infrastructure Control and Compliance,* International Society for Pharmaceutical Engineering (ISPE), Second Edition, August 2017, www.ispe.org.

112 . NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0), National Institutes of Standards and Technology (NIST), January 2023, doi.org/10.6028/NIST.AI.100-1, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

113. EudraLex Volume 4 – Guidelines for Good Manufacturing Practices for Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems, June 2011, https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en.

114. "Regulatory considerations on artificial intelligence for health," World Health Organization (WHO), 2023, www.who.int.

115. FDA/Health Canada/MHRA Good Machine Learning Practice for Medical Device Development: Guiding Principles, October 2021, US Food and Drug Administration (FDA), www.fda.gov, Health Canada, www.canada.ca/en/health-canada.html, United Kingdom Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency.

116. Altrabsheh, E., Heitmann, M., Steinmüller, P., Pastori Vinco, B., "The Road to Explainable AI in GXP-Regulated Areas," *Pharmaceutical Engineering,* January/February 2023, Vol. 43, No. 1, pp. 24–32, www.ispe.org.

117. United Nations (UN) Sustainable Development Goals, 2015, https://sdgs.un.org/goals.

118. Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC, 29 June 2023, https://eur-lex.europa.eu/eli/reg/2023/1230/oj.

119. Stockton, B.M., Staib, E.J., Shaw, D., Donald, J.C., "The Machine Learning Life Cycle (MLLC): Key Artifacts, Considerations, and Questions," iSpeak Blog, International Society for Pharmaceutical Engineering (ISPE), 14 March 2024, www.ispe.org/pharmaceutical-engineering/ispeak/machine-learning-life-cycle-mllc-key-artifacts-considerations-and.

120. Digital Health and Artificial Intelligence Glossary – Educational Resource, US Food and Drug Administration (FDA), 26 September 2024, www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource.

121. "What is semi-supervised learning?" IBM Topics, 12 December 2023, www.ibm.com/topics/semi-supervised-learning.

122. Vaswani, A., et. al., "Attention Is All You Need," arXiv, Cornell University, 12 June 2017, Revised 2 August 2023 (v7), https://arxiv.org/abs/1706.03762.

123. Phillips, P. J., et. al., NIST Interagency/Internal Report 8312: "Four Principles of Explainable Artificial Intelligence," 29 September 2021, National Institute of Standards and Technology (NIST), doi.org/10.6028/NIST.IR.8312, https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf.

124. Broniatowski, D.A., "Psychological Foundations of Explainability and Interpretability in Artificial Intelligence," NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, 12 April 2021, www.nist.gov/publications/psychological-foundations-explainability-and-interpretability-artificial-intelligence.

125. Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No. 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act), 20 November 2024, https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng.

126. ISO/IEC TR 27563:2023 Security and privacy in artificial intelligence use cases — Best practices, International Organization for Standardization (ISO), www.iso.org, and International Electrotechnical Commission (IEC), www.iec.ch.

127. FDA Guidance for Industry and Food and Drug Administration Staff: Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions, US Food and Drug Administration (FDA), 27 September 2023, www.fda.gov/media/119933/download.

128. FDA Guidance for Industry and Food and Drug Administration Staff: Postmarket Management of Cybersecurity in Medical Devices, US Food and Drug Administration (FDA), 28 December 2016, www.fda.gov/media/95862/download.

129. AICPA SOC 2® – SOC for Service Organizations: Trust Services Criteria, Association of International Certified Professional Accountants (AICPA), www.aicpa-cima.com.

130. "Introduction to the HITRUST CSF, Version 11.4.0," HITRUST Alliance, December 2024, https://hitrustalliance.net/.

131. Cybersecurity Infrastructure Agency (CISA), www.cisa.gov.

132. Artificial Intelligence Program: Research on AI/ML-Based Medical Devices, US Food and Drug Administration (FDA), 26 September 2024, www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/artificial-intelligence-program-research-aiml-based-medical-devices.

133. FDA/Health Canada/MHRA Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices, October 2023, US Food and Drug Administration (FDA), www.fda.gov, Health Canada, www.canada.ca/en/health-canada.html, United Kingdom Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency.

134. ISO, ISO/IEC TR 29119 Software and systems engineering – Software testing, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

135. Classify Your Medical Device, US Food and Drug Administration (FDA), 7 February 2020, www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device.

136. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, 5 May 2017, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32017R0745.

137. 21 CFR Part 820 – Medical Devices, Quality System Regulation, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

138. AAMI TIR57:2016/(R)2023 Principles for Medical Device Security - Risk Management, Association for the Advancement of Medical Instrumentation (AAMI), webstore.ansi.org/standards/AAMI/aamitir572016r2023.

139. IEC 62366-1:2015 Medical devices Part 1: Application of usability engineering to medical devices, International Electrotechnical Commission (IEC), www.iec.ch.

140. eSTAR Program, US Food and Drug Administration (FDA), US Food and Drug Administration (FDA), 28 May 2025, www.fda.gov/medical-devices/how-study-and-market-your-device/estar-program.

141. Total Product Life Cycle for Medical Devices, US Food and Drug Administration (FDA), 6 September 2023, www.fda.gov/about-fda/cdrh-transparency/total-product-life-cycle-medical-devices.

142. FDA Executive Summary for the Digital Health Advisory Committee Meeting: Total Product Lifecycle Considerations for Generative AI-Enabled Devices, US Food and Drug Administration (FDA), 20–21 November 2024, www.fda.gov/media/182871/download.

143. ISO/TR 24971:2020 Medical devices — Guidance on the application of ISO 14971, International Organization for Standardization (ISO), www.iso.org.

144. IG-NB Questionnaire “Artificial Intelligence (AI) in medical devices”, Version 5.1, German Notified Bodies (IG-NB), 4 March 2024, www.ig-nb.de.

145. 21 CFR Part 814 – Premarket Approval of Medical Devices, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

146. FDA Guidance for Industry: Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software, US Food and Drug Administration (FDA), 14 January 2005, www.fda.gov/media/72154/download?attachment.

147. FDA Draft Guidance for Industry and Food and Drug Administration Staff: Select Updates for the Premarket Cybersecurity Guidance: Section 524B of the FD&C Act, US Food and Drug Administration (FDA), 13 March 2024, www.fda.gov/media/176944/download.

148. NIST SP 800-30 Revision 1: Guide for Conducting Risk Assessments, National Institutes of Standards and Technology (NIST), September 2012, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf.

149. European Commission MDCG 2019-16 Guidance on Cybersecurity for medical devices, European Commission Medical Device Coordination Group (MDCG), December 2019, https://ec.europa.eu/docsroom/documents/41863.

150. Principles and Practices for Medical Device Cybersecurity, International Medical Device Regulators Forum (IMDRF), 18 March 2020, www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-200318-pp-mdc-n60.pdf.

151. Team-NB Position Paper: Cyber Security, The European Association of Medical devices Notified Bodies (TEAM-NB), 5 October 2022, www.team-nb.org/wp-content/uploads/members/M2022/Team-NB-PositionPaper-CyberSecurity-V1-20221005.pdf.

152. IG-NB Questionnaire “Cybersecurity for Medical Devices - Technical Documentation,” German Notified Bodies Alliance (IG-NB), 21 March 2023, www.ig-nb.de/fileadmin/user_upload/ig-nb/Questionnaire_Cybersecurity_for_Medical_Devices_-_Technical_Documentation_-_Version_1.pdf.

153. IG-NB Questionnaire “Cybersecurity for Medical Devices - Audit,” German Notified Body Alliance (IG-NB), 21 March 2023, www.ig-nb.de/fileadmin/user_upload/ig-nb/Questionnaire_Cybersecurity_for_Medical_Devices_-_Audit_-_Version_1.pdf.

154. IEC 81001-5-1:2021 Health software and health IT systems safety, effectiveness and security, Part 5-1: Security — Activities in the product life cycle, International Electrotechnical Commission (IEC), www.iec.ch

155. FDA Discussion Paper and Request for Feedback: Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products, US Food and Drug Administration (FDA), May 2023 (Revised February 2025), www.fda.gov/media/167973/download?attachment.

156. FDA Guidance for Industry and Food and Administration Staff: Deciding When to Submit a 510(k) for a Software Change to an Existing Device, US Food and Drug Administration (FDA), 25 October 2017, www.fda.gov/media/99785/download.

157. National framework for the assurance of artificial intelligence in government, Therapeutic Goods Administration (TGA), 21 June 2024, www.finance.gov.au/government/public-data/data-and-digital-ministers-meeting/national-framework-assurance-artificial-intelligence-government.

158. Understanding regulation of software Based Medical Devices, Therapeutic Goods Administration (TGA), 23 September 2024, www.tga.gov.au/how-we-regulate/manufacturing/medical-devices/manufacturer-guidance-specific-types-medical-devices/regulation-software-based-medical-devices#guidance.

159. MHRA Guidance: Software and Artificial Intelligence as a Medical Device, Medicines & Healthcare products Regulatory Agency (MHRA), 14 June 2023, www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap.

160. ISO 9000 Quality management system, International Organization for Standardization (ISO), www.iso.org.

161. "A pro-innovative approach to AI regulation," Office for Artificial Intelligence, Department for Science, Innovation & Technology, UK Government, 3 August 2023, www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.

162. FDA Draft Guidance for Industry: Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products, US Food and Drug Administration (FDA), February 2023, www.fda.gov/media/164960/download.

163. Food and Drug Omnibus Reform Act (FDORA) 2022, US Food and Drug Administration (FDA), 29 December 2022, www.fda.gov/regulatory-information/selected-amendments-fdc-act/food-and-drug-omnibus-reform-act-fdora-2022.

164. EMA Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data sources, Version 2.0, European Medicines Agency (EMA), 8 April 2025 www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf.

165. Forrest, S., "Artificial Intelligence/ Machine Learning (AI/ ML)-Enabled Medical Devices: Tailoring a Regulatory Framework to Encourage Responsible Innovation in AI/ML," Center for Devices & Radiological Health (CDRH), US Food and Drug Administration (FDA), Accessed 25 June 2025, www.fda.gov/media/160125/download.

166. EMA Concept Paper on the revision of Annex 11 of the guidelines on Good Manufacturing Practice for medicinal products – Computerized Systems, European Medicines Agency (EMA), 19 September 2022, www.ema.europa.eu/en/documents/regulatory-procedural-guideline/concept-paper-revision-annex-11-guidelines-good-manufacturing-practice-medicinal-products-computerised-systems_en.pdf.

167. Artificial Intelligence/Machine Learning (AI/ML) – Based Software as a Medical Device (SaMD) Action Plan, US Food and Drug Administration (FDA), January 2021, www.fda.gov/media/145022/download .

168. Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, US White House, 3 December 2020 www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

169. White Paper: On Artificial Intelligence – A European approach to excellence and trust, European Commission, 19 February 2020, https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

170. FDA Discussion Paper and Request for Feedback: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) – Based Software as a Medical Device (SaMD), US Food and Drug Administration (FDA), 2 April 2019, www.fda.gov/media/122535/download?attachment.

171. Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence, European Commission, 8 April 2019, https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.

172. Framework for FDA's Real-World Evidence Program, US Food and Drug Administration (FDA), December 2018, www.fda.gov/media/120060/download.

173. "Definition of BLOCKCHAIN," www.merriam-webster.com, 15 July 2022, www.merriam-webster.com/dictionary.

174. ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology, International Organization for Standardization (ISO), www.iso.org and International Electrotechnical Commission (IEC), www.iec.ch.

175. Kumar, S., "Deep Learning," Python for Accounting and Finance, Palgrave Macmillan, 2024, pp. 459–500.

176. "Jaccard Similarity," *ScienceDirect,* 2016, www.sciencedirect.com/topics/computer-science/jaccard-similarity.

177. "What are large language models (LLMs)?" IBM Topics, 2 November 2023, www.ibm.com/think/topics/large-language-models.

178. Gross, R., "Psychology: The Science of Mind and Behaviour," Eighth Edition, Hodder Education, 2020.

179. "Levenshtein distance (definition)," National Institutes of Standards and Technology (NIST), 2019, https://xlinux.nist.gov/dads/HTML/Levenshtein.html

180. "Definition of RED TEAM," National Institutes of Standards and Technology (NIST), https://csrc.nist.gov/glossary/term/Red_Team.

181. National Institutes of Standards and Technology (NIST), www.nist.gov.

182. IEEE Standards Collection, Software Engineering, Institute of Electrical and Electronics Engineers, 1994.

183. Erdmann, N., Blumenthal, R., Baumann, I., Kaufmann, M., "AI Maturity Model for GxP Application: A Foundation for AI Validation," *Pharmaceutical Engineering,* March/April 2022, Vol. 42, No. 2, pp. 18-24, www.ispe.org.

31 Appendix G2 – Glossary

31.1 Acronyms and Abbreviations

AAMI	Association for the Advancement of Medical Instrumentation
AI	Artificial Intelligence
AICPA	Association of International Certified Professional Accountants
ALCOA++	Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, Available, Traceable
ANOVA	Analysis of Variance
API	Application Programming Interface
APQC	American Productivity & Quality Center
ASTM	American Society for Testing and Materials
BLEU	Bilingual Evaluation Understudy
BSI	Federal Office of Information Security (Germany)
CAPA	Corrective and Preventive Action
CBOM	Cyber Bill of Materials
CMMI	Capability Maturity Model Integration
CNN	Convolutional Neural Networks
CPP	Critical Process Parameter
CQA	Critical Quality Attribute
CRISP-ML(Q)	Cross-Industry Standard Process for the development of Machine Learning applications with Quality assurance methodology
DHF	Design History File
EDA	Exploratory Data Analysis
EMA	European Medicines Agency (EU)
ETL	Extract, Transform, Load
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FD&C	Food, Drug, and Cosmetic (US)
FDA	Food and Drug Administration (US)
FDORA	Food and Drug Omnibus Reform Act (US)
FMEA	Failure Mode and Effects Analysis
GAN	Generative Adversarial Network
GCP	Good Clinical Practice
GDP	Good Distribution Practice
GDPR	General Data Protection Regulation (EU)

GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
GPU	Graphics Processing Unit
GQP	Good Quality Practice
GVP	Good Pharmacovigilance Practice
GxP	Good "x" Practice
HACCP	Hazard Analysis and Crucial Control Points
HAZOP	Hazard Operability Analysis
HIPAA	Health Insurance Portability and Accountability Act (US)
HN	Hopfield Network
HSE	Health, Safety, and Environment
IaaS	Infrastructure as a Service
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics
IG-NB	German Notified Bodies Alliance
IMDRF	International Medical Device Regulators Forum
IoT	Internet of Things
IP	Intellectual Property
ISA	International Society of Automation
ISO	International Organization for Standardization
IT	Information Technology
KEDB	Known Error Database
KNN	K-Nearest Neighbor
KPI	Key Performance Indicator
LIME	Local Interpretable Model-agnostic Explanations
LIMS	Laboratory Information Management System
LLM	Large Language Model
LSTM	Long Short-Term Memory
MDCG	Medical Device Coordination Group (US)
MHLW	Ministry of Health, Labour and Welfare (Japan)
MHRA	Medicines & Healthcare products Regulatory Agency (UK)
ML	Machine Learning
MLOps	Machine Learning Operations

MLP	Multilayer Perceptron
NIST	National Institutes of Standards and Technology
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
OLA	Operational Level Agreement
OTS	Off-the-Shelf
PaaS	Platform as a Service
PCA	Principal Component Analysis
PCCP	Predetermined Change Control Plan
PET	Privacy-Enhancing Technologies
PHA	Preliminary Hazard Analysis
PHI	Personal Health Information
PHS	Public Health Service (US)
PII	Personally Identifiable Information
PMA	Premarket Approval
PMS	Post-Market Surveillance
PoC	Proof of Concept
QA	Quality Assurance
QbD	Quality by Design
QMS	Quality Management System
QRM	Quality Risk Management
RAG	Retrieval Augmented Generation
RAMM	Risk Analysis and Mitigation Matrix
RBAC	Role-Based Access Control
RCA	Root Cause Analysis
RDI	Records and Data Integrity
RNN	Recurrent Neural Network
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
RPO	Recovery Point Objectives
RTO	Recovery Time Objectives
SaaS	Software as a Service
SaMD	Software as Medical Devices
SBOM	Software Bill of Materials
SHA	Software Hazard Analysis
SHAP	SHapley Additive exPlanations
SiMD	Software in Medical Devices

SLA	Service Level Agreement
SMART	Specific, Measurable, Achievable, Relevant, Time-bound
SME	Subject Matter Expert
SOC	Systems and Organization Controls
SOP	Standard Operating Procedure
SOUP	Software of Unknown Provenance
SOX	Sarbanes-Oxley
TEAM-NB	The European Association of Medical devices Notified Bodies
TEVV	Test, Evaluation, Validation and Verification (NIST)
TGA	Therapeutic Goods Administration (Australia)
UK	United Kingdom
US	United States
VAE	Variational Autoencoder
XAI	Explainable AI

31.2 Definitions

AI Sub-system Adaptiveness

The ability to automatically perform updates and thereby automatically self-improve.

AI Use Case

A use case of automation that relies on the use of AI.

AI Function

An AI function is a function of a computer system that includes at least one AI sub-system as one of its modules.

AI Sub-system

A module that includes the use of AI. It includes a model or an ensemble of models and may include pre- and post-processing functionality.

AI System

An AI system is a computer system that includes the use of AI.

Relying on Grobelnik, Perset, and Russell [42], a common definition describing such a system is the following, which implicitly defines AI:

> *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."*

AI-enabled Computerized System

A computerized system that includes the use of AI, i.e., it contains at least one AI System.

AI-enabled Software Products

A software product provided by a supplier that includes the use of AI.

AI Literacy

AI literacy includes foundational understanding of how AI methods function, the considerations required to activate such innovation effectively, and their implications within the context of use.

'AI literacy' means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause. [24]

AI Project

A project concerned with the adoption, or change of an AI-enabled computerized system, performing activities in the project phase of the AI-enabled computerized system life cycle.

Algorithm

A finite set of instructions that can be executed by a computer.

AI Governance

The sum of arrangements at an organizational level, including rules, structures, and processes, to ensure that AI-enabled systems are safe and adhere to ethical principles.

AI Sub-system Autonomy

The degree by which the system relies on embedded controls rather than human verification of functions that impact patient safety, product quality, and data integrity.

Blockchain [173]

A digital database containing information (such as records of financial transactions) that can be simultaneously used and shared within a large decentralized, publicly accessible network.

Case Data Set

The totality of data used for AI sub-system development purposes.

Confusion Matrix (ISO [134])

Table used to describe the performance of a classifier on a set of test data for which the true and false values are known.

Context of Use (FDA [55])

The context of use defines the specific role and scope of the AI model used to address a question of interest.

Control Strategy (ICH Q10 [35]

A planned set of controls, derived from current product and process understanding, that assures process performance and product quality.

Cost Function

An objective function that is designed to evaluate the output of a machine learning model to serve as optimization criterion in determining or iterating parameters of the machine learning model.

Data Card (FDA [120])

A structured report of relevant characteristics of datasets needed by stakeholders for AI development and evaluation. It contains a descriptive section including descriptive information such as number of samples, collection protocols and associated metadata, and a scorecard section, a quantitative analysis reporting dataset characteristics using relevant criteria and metrics.

Data Fitness for Purpose

Data that is reliable, relevant, representative, and abundant in the context of use.

Data Pipeline

A data "pipeline" refers to a comprehensive and organized process or workflow that encompasses (possibly) multiple stages or steps for data processing and transformation. It involves the flow of data through a series of interconnected stages or components, often automated, to perform specific operations such as ETL, data cleansing, analysis, and ultimately delivering the processed data to its intended destination or purpose.

Data Quality Dimensions

Desirable properties of data, or a data set; they support reliable data as one element of fit for purpose data.

Data Science [54]

Data science is an interdisciplinary field that involves using scientific methods, algorithms, and statistical models to extract insights and knowledge from data. Data science combines elements of computer science, statistics, and domain expertise to make sense of large, complex data sets.

Data Set (ISO [174])

Collection of data with a shared format.

Deep Learning

An approach to creating rich hierarchical representations through the training of neural networks with one or more hidden layers. (ISO [134])

Deep learning is a subfield of machine learning that focuses on developing algorithms and models inspired by the structure and function of the human brain.

The term "deep" refers to the fact that deep learning models typically consist of multiple layers of interconnected artificial neurons. [175]

Dynamic System

A dynamic AI-enabled computerized system (or in short dynamic system) is an AI-enabled computer system in which AI sub-systems may evolve in an automated way without approval by a human operator. This may include fully automatized retraining, model quality assurance and model deployment procedures up to incremental learning systems by which model parameters may change with any new observations.

Explainability [33]

The degree to which a basis for a decision or action can be explained or how an output or result was reached, in a way that a person can understand.

Foundation Models

Pre-trained models that provide general capabilities and can be used as is or fine-tuned via learning on own data to tailor the model to a specific use case.

Generative AI

AI that can create human-like content like texts, images, videos, or other data modalities.

Ground Truth

The actual outcome of data that a model seeks to predict.

GxP Regulation [2]

The underlying international pharmaceutical requirements, such as those set forth in the US FD&C Act, US PHS Act, FDA regulations, EU Directives, Japanese regulations, or other applicable national legislation or regulations under which a company operates. These include but are not limited to:

- Good Manufacturing Practice (GMP) (pharmaceutical, including Active Pharmaceutical Ingredient (API), veterinary, and blood)
- Good Clinical Practice (GCP)
- Good Laboratory Practice (GLP)
- Good Distribution Practice (GDP)
- Good Quality Practice (GQP)
- Good Pharmacovigilance Practice (GVP)
- Medical Device Regulations
- Prescription Drug Marketing Act (PDMA)

Harm (ICH [31])

Damage to health, including the damage that can occur from loss of product quality or availability.

Hazard (ICH [31])

The potential source of harm.

Hyperparameter (ISO [174])

Characteristic of a machine learning algorithm that affects its learning process.

Incident [2]

Operational event which is not part of standard operation.

Interpretability [124]

The ability to contextualize a model's output in a manner that relates it to the system's designed functional purpose, and the goals, values, and preferences of end users.

Iterative Experimentation

A process to subsequently derive models including pre- and postprocessing functionality, aiming to achieve a best-performing model given one or a set of model performance indicators.

Jaccard Similarity [176]

Jaccard similarity refers to a measure of similarity between two sets of keywords. It is used to determine if clusters containing raw messages are duplicates by computing the similarity score between the sets of keywords associated with each cluster.

Large Language Models [177]

Large language models (LLMs) are a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks; see also Generative AI.

Learning

Learning is the process of acquiring new understanding, knowledge, behaviors, skills, values, attitudes, and preferences. [178]

In an AI context, learning is the process of a machine or computer system improving its performance on any actions/decisions by "learning" from data. This involves developing algorithms or models that can analyze data and discover patterns, relationships, and rules that can be used to make predictions, classify new data, or take actions based on that data.

There are several learning strategies in ML, including:

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning
- Transfer Learning

Levenshtein Distance [179]

The smallest number of insertions, deletions, and substitutions required to change one string or tree into another.

Machine Learning (ML) [48]

Machine learning is a type of artificial intelligence that allows machines to learn from data without being explicitly programmed.

Machine Learning Operations (MLOps)

A collection of techniques and approaches to support activities regarding data and model management in the context of AI-enabled systems.

ML sub-system

An artificial intelligence sub-system that leverages machine learning.

Model

A model is a complex rule-based algorithm, or mathematical algorithms with parameters (weights) arranged in an architecture that allows training that represents a certain aspect of a real-world process.

A model is seen as a sub-program, hence part of an AI sub-system.

Model Card

A structured report of relevant technical characteristics of an AI model and benchmark evaluation results in a variety of conditions, such as across different cultural, demographic, or phenotypic groups and intersectional groups that are relevant to the intended application domains. Model cards also provide information about the context in which models are intended to be used and details of how their performance was assessed.

Model Design Space

A description of applicable learning methods and learning types for a given use case. The model design space provides an overview of model development strategies that should be examined during the iterative model engineering and evaluation cycle.

Model Development Life Cycle

A development approach that includes the definition of a model design space to determine the scope of applicable models, conduct of iterative experimentation to derive models, and selection of best-performing models.

Model Release Candidate

A model release candidate is a version of a model that is considered ready for final testing, evaluation, and potential release according to its measured performance metrics. Often they are selected by comparing its performance metrics against other model candidates.

Natural Language Processing (NLP)

Discipline concerned with the way systems acquire, process, and interpret language that is or was in active use in a community of people and whose rules are deduced from usage. (ISO [174])

A subfield of AI and linguistics that enables computers to understand, process, interpret, and generate human language. NLP systems can perform tasks such as text classification, sentiment analysis, and translation, using techniques from computational linguistics and ML to process and analyze natural language data. (FDA [120])

Non-AI-enabled Computerized System

A computerized system that does not make use of AI, i.e., that does not contain an AI System.

Operational Level Agreement

An agreement that outlines how the organization will interact to achieve the objectives or goals of the SLA. OLAs track service commitments like response times and incident resolutions. For the AI sub-system, the OLA shall specify how quickly the respective team shall respond when the model encounters an anomaly depending on the severity.

Polyglot Data Storage

Data storage comprises structured and unstructured data.

Proof of Concept (PoC)

Proof of concept is a conceptual activity to demonstrate the feasibility and the potential of a solution approach. It includes considerations related to the problem statement and use case definition, regulatory requirements, data, and risks. A Proof of Concept may include implementing a prototype to demonstrate the capability of AI technology to solve the chosen problem. The role of the PoC is to enable decision makers to answer questions while maximizing value and minimizing risk.

Process Owner [2]

This is the owner of the business process or processes being managed. The process owner is ultimately responsible for ensuring that the computerized system and its operation is in compliance and fit for intended use in accordance with applicable company Standard Operating Procedures (SOPs). The process owner may also be the system owner. The process owner may be the *de facto* owner of the data residing on the system (data owner) and therefore, ultimately responsible for the integrity of the data. Process owners are typically the head of the functional unit using the system.

(cf. System Owner)

Quality by Design (QbD) (ICH [84])

A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management.

Quality Risk Management (QRM) (ICH [31])

A systematic process for the assessment, control, communication and review of risks to the quality of the drug (medicinal) product across the product lifecycle.

Quality System (ICH [31])

The sum of all aspects of a system that implements quality policy and ensures that quality objectives are met.

Red Team [180]

A group of people authorized and organized to emulate a potential adversary's attack or exploitation capabilities against an enterprise's security posture. The Red Team's objective is to improve enterprise cybersecurity by demonstrating the impacts of successful attacks and by demonstrating what works for the defenders (i.e., the Blue Team) in an operational environment. Also known as Cyber Red Team.

Reinforcement Learning (FDA [120])

A ML approach where a model (or agent) learns by taking actions and getting rewards or penalties through its interactions with an environment. The model learns from the consequences of its actions, rather than from being explicitly taught, and selects its actions based on its past experiences (exploitation) and by making new choices (exploration), which is essentially trial and error learning.

Requirement [160]

Need or expectation that is stated, generally implied or obligatory.

Risk (ICH [31])

The combination of the probability of occurrence of harm and the severity of that harm.

Risk Analysis (ICH [31])

The estimation of risk associated with the identified hazards.

Risk Assessment (ICH [31])

A systematic process of organizing information to support a risk decision to be made within a risk management process. It consists of the identification of hazards and the analysis and evaluation of risks associated with exposure to those hazards.

Risk Communication (ICH [31])

The sharing of information about risk and risk management between the decision maker and other stakeholders.

Semi-supervised Learning (FDA [120])

ML algorithms that leverage both unsupervised and supervised techniques. Supervised learning techniques are trained using labeled data, while unsupervised learning techniques are trained using unlabeled data. Labeling or annotation is the process of attaching descriptive information to data. Data itself are unchanged in the annotation process.

Service Level Agreement (SLA)

An agreement between the service provider and the customer that defines the performance expectations expected to be provided for the AI sub-system. The SLA shall define the scope of the service provider, performance targets the service provider will achieve, including response times and resolution times, and the responsibilities of each of the roles.

Severity (ICH [31])

A measure of the possible consequences of a hazard.

Software Life Cycle [181]

Period beginning when a software product is conceived and ending when the product is no longer available for use. The software life cycle is typically broken into phases denoting activities such as requirements, design, programming, testing, installation, and operation and maintenance.

Source Code [52]

1. Computer instructions and data definition expressed in a form suitable for input to an assembler, compiler or other translator. (IEEE)

2. The human readable version of the list of instructions [program] that causes a computer to perform a task.

Specification (IEEE [182])

A document that specifies, in a complete, precise, verifiable manner, the requirements, design, behavior, or other characteristics of a system or component, and often, the procedures for determining whether these provisions have been satisfied.

Note: Specifications may also be captured in other forms than documents.

Subject Matter Expert (SME) (ASTM [16])

Those individuals with specific expertise in a particular area or field. Subject Matter Experts should take the lead role in the verification of computerized systems. Subject Matter Expert responsibilities include planning and defining verification strategies, defining acceptance criteria, selection of appropriate test methods, execution of verification tests, and reviewing results.

Supervised Learning (FDA [120])

ML algorithms where labeled data is provided, and algorithms are trained using the labeled data. Labeling or annotation is the process of attaching descriptive information to data. Data itself is unchanged in the annotation process.

Supplier [2]

An organization or individual internal or external to the user associated with the supply and/or support of products or services at any phase throughout a system's life cycle.

System Owner [2]

The system owner is responsible for the availability, and support and maintenance, of a system, and for the security of the data residing on that system. The system owner is responsible for ensuring that the computerized system is supported and maintained in accordance with applicable company SOPs. The system owner also may be the process owner (e.g., for IT infrastructure systems or systems not directly supporting GxP). For systems supporting regulated processes and maintaining regulated data and records, the ownership of the data resides with the GxP-process owner, not the system owner.

The system owner acts on behalf of the users. The system owner for larger systems will typically be from IT or engineering functions. Global IT systems may have a global system owner and a local system owner to manage local implementation.

(cf. Process Owner)

Test Data Set (ISO [174])

Data used to assess the performance of a final model.

Test/Testing

Activity in which a system or component is executed under specified conditions, the results are observed or recorded, and an evaluation is made of some aspect of the system or component. [2]

Testing may include ad-hoc testing, error guessing, exploratory testing, and day in the life testing, as well as other test types. (FDA [120])

Testing, functional (IEEE [182])

1. Testing that ignores the internal mechanism or structure of a system or component and focuses on the outputs generated in response to selected inputs and execution conditions.

2. Testing conducted to evaluate the compliance of a system or component with specified functional requirements and corresponding predicted results.

Synonyms: black-box testing, input/output driven testing.

Testbed [46]

A facility or mechanism equipped for conducting rigorous, transparent, and replicable testing of tools and technologies, including AI and PETs [Privacy-Enhancing Technologies], to help evaluate the functionality, usability, and performance of those tools or technologies.

Test Case (ISO [134])

A set of preconditions, inputs (including actions, where applicable) and expected results, developed to drive the execution of a test item to meet test objectives.

Test Plan (ISO [134])

Detailed description of test objectives to be achieved and the means and schedule for achieving them, organized to coordinate testing activities for some test item or set of test items.

Training Epochs

The number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training data set. Too few epochs can result in an underfit model, whereas too many epochs can lead to overfitting.

Training Data Set (ISO [174])

Data used to train a machine learning model.

Unsupervised Learning (FDA [120])

ML algorithms that only make use of unlabeled data during training. Unsupervised learning seeks to uncover hidden patterns or structures within the data.

User [2]

The pharmaceutical customer or user organization contracting a supplier to provide a product. In the context of this document it is, therefore, not intended to apply only to individuals who use the system, and is synonymous with customer.

Validation [52]

Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes.

Validation Data Set (FDA [120])

The fraction of the case data set that is used to evaluate the model's performance during iterative experimentation.

Also called tuning data (set) in contexts where confusion may arise with other interpretations of validation data as part of regulatory submissions.

Verification

1. Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled. [160]

2. A systematic approach to verify that manufacturing systems, acting singly or in combination, are fit for intended use, have been properly installed, and are operating correctly. This is an umbrella term that encompasses all types of approaches to assuring systems are fit for use such as qualification, commissioning and qualification, verification, system validation, or other. [16]

3. In general, the demonstration of consistency, completeness, and correctness of the software at each stage and between each stage of the development life cycle.

Watermarking (FDA [120])

The process of embedding an identifying pattern in a piece of media in order to track its origin —including into outputs such as images, audio, video, and digital text—for the purposes of verifying the authenticity of the output or the identity or characteristics of its provenance, modifications, or conveyance.

XAI Method

A method used to provide insights on how a model has derived the model output from model input; see NISTIR 8312. [123]

3001 N. Rocky Point Dr. E., Suite 200-242, Tampa, Florida 33607 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

www.ISPE.org