# MASTER THESIS | MASTER'S THESIS

Titel | Title

Application of Artificial Intelligence (AI) in the Context of Visual Inspection of Parenterals: Opportunities, Risks and Regulatory Framework in the GMP Environment

verfasst von | submitted by

Mag.pharm. Patrick Gassenbauer

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

| | |
|---|---|
| Studienkennzahl lt. Studienblatt \| Degree programme code as it appears on the student record sheet: | UA 992 580 |
| Universitätslehrgang lt. Studienblatt \| Postgraduate programme as it appears on the student record sheet: | Pharmazeutisches Qualitätsmanagement MSc |
| Betreut von \| Supervisor: | Dipl.-Ing. (FH) Dr. Michael Zwitkovits MSc |

# Abstract (DE)

Diese Arbeit untersucht, wie Künstliche Intelligenz und Maschinelles Lernen (KI/ML) die visuelle Inspektion (VI) von Parenteralia in GMP-regulierten Herstellprozessen verbessern können, indem sie Limitationen der manuellen visuellen Inspektion (MVI) sowie der regelbasierten automatisierten visuellen Inspektion (AVI) adressieren. Vier Forschungsfragen strukturieren die Arbeit:

- (RQ1) Welche potenziellen Anwendungsfälle für KI/ML-Software gibt es in der visuellen Sichtkontrolle von Parenteralia?
- (RQ2) Welche Chancen und Herausforderungen bringt KI/ML im GMP-Bereich mit sich?
- (RQ3) Deckt der aktuelle Rechtsrahmen in Europa und den USA die Bedürfnisse der Pharmaindustrie im Sinne von verbindlichen KI/ML-Leitlinien ab?
- (RQ4) Welche Best Practices für KI/ML lassen sich aus der Medizinprodukteindustrie ableiten?

Die Auswertung von Peer Reviews und grauer Literatur zeigt, dass Deep-Learning-Ansätze, insbesondere Convolutional Neural Networks (CNN), Inspektionsfehler reduzieren, die Entwicklung von AVI-Vision-Recipes beschleunigen und datenbasierte Prozesseinblicke ermöglichen können. Voraussetzung ist, dass robustes Daten-Governance und ein wirksames Qualitätsrisikomanagement bestehen. Guideline Entwürfe wie der neue EU-GMP-Annex 22 zu Künstlicher Intelligenz sowie US-amerikanische FDA-Dokumente beginnen, bestehende Regulierungslücken zu schließen; dennoch werden weiterhin klare Strategien für die Validierung, das Change Control sowie die Erklärbarkeit von KI/ML-Modellen für pharmazeutische Hersteller benötigt.

Die Arbeit kommt zum Schluss, dass der Einsatz von KI/ML-Anwendungen bei der visuellen Sichtkontrolle von Parenteralia das Potenzial hat, Inspektionsfehler und Kosten zu reduzieren sowie die Prozesseffizienz zu steigern. Für eine erfolgreiche Implementierung im GMP-Bereich ist ein umfassender, risikobasierter Ansatz erforderlich, der technische, regulatorische und operative Herausforderungen adressiert.

**Schlagwörter**: Visuelle Sichtkontrolle, Künstliche Intelligenz, Maschinelles Lernen, Deep Learning, Parenteralia, Arzneimittel, Good Manufacturing Practice, GMP, Leitlinie, EMA, FDA, Medizinprodukt, EU GMP Annex 22

# Abstract (EN)

This thesis investigates how artificial intelligence and machine learning (AI/ML) can enhance visual inspection (VI) of parenteral drug products in GMP-regulated manufacturing by overcoming limitations of manual visual inspection (MVI) and rule-based automated visual inspection (AVI). Four research questions structure the work:

- (RQ1) What are potential use cases of AI/ML software in VI of parenteral drug products?
- (RQ2) What challenges and opportunities are associated with the adoption of AI/ML in GMP-regulated environments?
- (RQ3) Does the current legal framework in Europe and the United States of America meet the need of the pharmaceutical industry for AI/ML guidance in GMP?
- (RQ4) What AI/ML best practices can be derived from the medical device industry?

The assessment of peer-reviewed and grey literature shows that deep-learning (DL) approaches, particularly convolutional neural networks (CNN), can reduce inspection errors, accelerate AVI recipe development, and yield data-driven process insights, provided robust data governance and risk management are in place. Emerging documents, such as the new EU GMP Annex 22 on Artificial Intelligence and draft FDA guidance, begin to close regulatory gaps; however, clear strategies for the validation, change control, and explainability of AI/ML models for pharmaceutical manufacturers are still needed.

The thesis concludes that AI/ML applications in VI of parenteral drug products holds considerable promise for reducing inspection errors, cost and enhancing process efficiency. Successful integration within GMP environments requires a comprehensive, risk-based approach that addresses technical, regulatory, and operational challenges.

**Keywords**: Visual Inspection, Artificial Intelligence, Machine Learning, Deep Learning, Regulatory Framework, Parenteral, Drug Product, Good Manufacturing Practice, GMP, Guideline, EMA, FDA, Medical Device, EU GMP Annex 22

# Acknowledgment

I would first like to express my deepest appreciation to my partner Laura-Elina, whose unwavering support, patience, and encouragement have been essential throughout this journey. Your belief in me, understanding during challenging moments, and motivation provided me with strength and determination to complete this work. I am truly thankful for your presence in my life.

I also extend my heartfelt appreciation to my family for their continuous love, patience, and encouragement.

Finally, I am profoundly grateful to my supervisor, Dr. Michael Zwitkovits, for his guidance, insightful advice, and continuous support throughout my master's thesis. His extensive knowledge, patience, and constructive feedback significantly contributed to my academic growth.

# Table of contents

# 1 Introduction

The pharmaceutical manufacturing industry is currently witnessing a paradigm shift driven by the integration of advanced technologies, most prominently artificial intelligence and machine learning (AI/ML). Visual inspection (VI) is a critical quality control process in the production of parenteral drug products, ensuring that both the products and their primary packaging meet stringent Good Manufacturing Practice (GMP) standards. Traditionally, this inspection has relied on manual visual – (MVI) and rule-based automated visual inspection (AVI) systems that, despite their long-standing use, are limited by human variability and the inflexibility of engineered algorithms when it comes to difficult to inspect products (DIPs). Consequently, there is an increasing motivation to leverage AI/ML techniques to enhance the robustness of visual inspection processes.

Recent developments in deep learning (DL) and computer vision (CV) have demonstrated promising capabilities in addressing the inherent challenges of traditional VI. AI/ML-driven methodologies, including convolutional neural networks (CNNs) and transfer learning, have shown potential in reducing inspection errors, both type 1 (false rejections) and type 2 (false acceptances), thereby improving batch yield, operational efficiency as well as product quality and patient safety. Nonetheless, the transition to AI/ML-enabled systems introduces a complex set of new challenges. These include risks related to data governance, the need for robust model validation, and the challenges associated with integrating these novel technologies within GMP environments.

This literature review addresses four research questions (RQ1–RQ4) to tackle the subject's complexity:

- **RQ1:** What are potential use cases of AI/ML software in visual inspection of parenteral drug products?
- **RQ2:** What challenges and opportunities are associated with the adoption of AI/ML in GMP-regulated environments?
- **RQ3:** Does the current legal framework in Europe and the United States of America meet the need of the pharmaceutical industry for AI/ML guidance in GMP?
- **RQ4:** What AI/ML best practices can be derived from the medical device industry?

This work describes the potential of AI/ML technologies to improve visual inspection processes and discusses the technical, regulatory, and operational challenges involved in integrating these technologies into pharmaceutical manufacturing. By reviewing the available literature across these four research questions, this work aims to give an overview on AI/ML use cases, implications when implementing it in GMP environments and potential regulatory expectations.

## 2 Background

Between 2017 and 2021, 33% of all FDA sterile injectable drug recall notices were due to visible particles (1) and related quality and safety concerns, resulting in several drug shortages (2,3). Since particles are not intended to be present in the drug product but technically unavoidable during the manufacturing process, they are viewed as contamination. While visible particles provide a good measure for GMP process control, the overall impact to patient safety is still inconclusive due assumptions based on limited observational data and animal studies as well as lack of clinical studies (2,4).

Other typical defects found in visual inspection are related to the product itself (gross over- or under-fill, discoloration, melt or collapsed lyophilization cakes), as well as to the container (cracks, chips, scratches, dirt on exterior) and the container-closure-system (leaking, missing or damaged components such as stoppers, seals, flip caps etc) (1). The 2023 PDA Visual Inspection Survey revealed that around half of the participating manufacturing sites (n = 172) still use manual visual inspection (MVI) for particles and container/closure inspections. Around 57% (n = 172 responses) plan to shift to automated visual inspection (AVI) with justifications such as increased capacity, improved quality and cost savings. When asked about future directions, the majority of responses indicate a move towards automated systems and the use of AI/ML (5).

Leading AVI equipment vendors such as Körber Pharma (6), Syntegon Technology (7), Brevetti (8), Stevanato Group (9), and other similar companies already offer systems with integrated AI software sub-systems. Hence it is to be expected that its usage in AVI and other use cases related to process monitoring and data analytics will emerge (10). However, to gain reliable data-driven insights from manufacturing processes using AI, the quality management system needs to provide robust data governance that ensures regulatory compliance to good documentation practices. Sound data must be the foundation of AI/ML models, since predictions are based on data used during model training. The quality of the AI solution is strongly determined by the quality of the data used for training the model.

The cautious and comparatively slow adoption of new technologies in pharmaceutical manufacturing is caused by several factors such as lack of regulatory clarity, technical challenges, and missing personnel with domain expertise (for example data scientists) related to innovative technologies (11,12). As a highly regulated industry, uncertainties due to adoption AI/ML that might affect product quality or patient safety naturally leads to careful and gradual implementation.

The following subsections will provide background information on topics of:



### 2.1 Regulatory framework

Many pharmaceutical companies produce drug products for the global market and manufacturers need to comply with GMP expectations of various regulatory authorities. For example, parenteral drug products intended for the European market need to adhere to EU GMP Annex 1 “Manufacture of Sterile Medicinal Products” which was recently revised and came into operation in August 2023 (13). Its section 8 contains requirements for inspection of all filled containers of parenteral drug products for extraneous contamination or other defects, classification of defects and their criticality, training qualification of personnel as well as trending and analyzing of data to detect unusual levels of inter-batch defects. If a manufacturer decides to use automated inspection systems, Annex 1 explicitly requires that this process needs be equal or better than manual inspection, de facto making MVI the gold standard that sets the bar for acceptance of new inspection technologies.

### 2.1.1 Computerized Systems

EU GMP Annex 11 defines a computerized system (CS) is a set of software and hardware components which together fulfill certain functionalities as part of GMP regulated activities (14). Since AVI machines use software for digital image processing, Annex 11 needs to be followed. Conventional software programming approaches follow a *deterministic* rule-based approach to decide whether the product is to be rejected or not. The "vision recipe" defines how raw images (input) are processed, which features correlate to specific defects and how the final decision (output) is made. Traditionally, humans decide which image features to extract, relying on vision engineers' expertise to identify and reject defective units. In contrast, AI is a paradigm shift from this traditional software development approach. This *probabilistic* software modeling approach can learn from existing data, identify patterns, and make predictions with a degree of autonomy; see **FIGURE 12**.

EU GMP Annex 11 came into operation in 2011, hence particulars of AI software are not covered. Other than the EU AI Act (15), no sector specific legislation or guideline is available for the use of AI in pharmaceutical manufacturing in EU. This may lead to uncertainty about regulatory expectations for AI/ML software, potentially slowing the adoption of AI/ML technologies in the pharmaceutical industry.

However, this gap is now being addressed in July 2025: the European Commission released draft revisions of EU GMP Chapter 4 (Documentation) and Annex 11 (Computerised Systems) alongside a new Annex 22 dedicated specifically to the use of AI in GMP environments.

### 2.1.2 Pharmacopeial requirements and recommendations

Pharmacopeias provide a compendium of monographs with either binding or recommendatory character. The most widely used and accepted pharmacopeias include methods for visual inspection of parenteral drug products. The European- (EP), US- (USP) and Japanese (JP) Pharmacopoeia achieved harmonization of their MVI methods, the International (INT) Pharmacopoeia describes a MVI-method that it is not intended for use by a manufacturer for batch release purposes but has almost identical requirements as the harmonized methods. **TABLE 1** gives an overview of compendial monographs that are relevant to visual inspection, excluding sub-visible particles, with either binding or non-binding characteristic (recommendations). **TABLE 2** compares compendial requirements and method parameters of the harmonized MVI methods as well as the INT method.

| Pharmacopeia | Chapter | Title | Binding? |
|---|---|---|---|
| EP | 2.9.20 | Particulate contamination : visible particles | Yes |
| EP | 5.17.2 | Recommendations on testing of particulate contamination : visible particles | No |
| EP | Dosage form monographs | - Monograph 0520: Parenteral preparations<br>- Monograph 1163: Eye preparations<br>- Monograph 2031: Monoclonal antibodies for human use<br>- Monograph 2811: Intravesical preparations | Yes |
| INTERN | 5.7 | Tests for particulate contamination | - |
| INTERN | 5.7.2 | Visible particles | - |
| INTERN | Dosage form monographs | Monograph 6.2.1.5: Parenteral preparations<br>Monograph 6.2.1.3: Ophthalmic preparations | - |
| JP | 6.06 | Foreign Insoluble Matter Test for Injections | Yes |
| JP | 6.07 | Insoluble Particulate Matter Test for Injections | Yes |
| JP | 6.08 | Insoluble Particulate Matter Test for Ophthalmic Solutions | Yes |
| USP | <1> | Injections and implanted drug products (parenterals)— product quality tests | Yes |
| USP | <1790> | Visual inspection of injections | No |
| USP | <771> | Ophthalmic products—quality tests | Yes |
| USP | <788> | Particulate matter in injections | Yes |
| USP | <789> | Particulate matter in ophthalmic solutions | Yes |
| USP | <790> | Visible particulates in injections | Yes |

*Table 1 Pharmacopeial monographs relevant to visual inspection with either binding or non-binding characteristic.*

| MVI Method parameter | USP < 790 > | EP 2.9.20 | JP 6.06 | INTERN 5.7.2* |
|---|---|---|---|---|
| **Illumination intensity** | 2000 – 3750 lux | 2000 – 3750 lux | 2000 – 3750 lux | 2000 – 3750 lux for clear glass ampoules |
| | | | 8000-10000 lux for aqueous injections in plastic containers | Higher values are preferable for colored glass and plastic containers |
| **Background** | Black and white background | | | |
| **Magnification** | No magnification | | | |
| **Inspection time** | Per unit: 5 s black+ 5 s white = 10 s | | | |
| **Acceptance criteria** | "Essentially free of visible particulates" | "Clear and practically free from particles" | "Free from readily detectable foreign insoluble matters" | Out of 20 containers: ≤ 1 container found with one or more particles |
| * The test is not intended for uses by a manufacturer for batch release purposes. | | | | |

*Table 2 Compendial requirements and method parameters for MVI*

While the explicit wording for the acceptance criteria differs ("essentially free of…", "…practically free from…" and "Free from readily detectable…"), all methods indirectly acknowledge the probabilistic nature of visual inspection. This important realization was first described in a publication by Julias Knapp and Harold Kusher in 1980 (16) and demonstrated that neither a human nor a robot can achieve a guaranteed 100% detection rate of visible particles even under optimal conditions. The statistical method, referred to as Knapp(-Kushner) methodology, is often used to establish a probability of detection (POD) for each defect type that is used as a performance benchmark for qualification of operators or equipment.

Compared to all other pharmacopoeias, recommendations in USP informational chapter <1790> provide a broad and holistic guidance on the topic of visual inspection. When comparing the scope of the recommendatory monographs of USP and EP, this approach is also reflected in the title:

- USP <1790> "Visual inspection of injections"
- EP 5.17.2 "Recommendations on testing of particulate contamination : visible particles"

The USP monograph does not limit the scope to *visible particular matter* but rather extends it to *visible inspection* that are critical to a comprehensive inspection process. Examples of this wider scope includes:

- Defects that may compromise the sterility of the product such as container-closure integrity defects (cracks, misplaced closures, or incomplete seals),
- Container system defects (visible defects in glass containers, elastomeric component, and aluminum seals),
- Other product characteristics (fill level, discoloration, clarity,…)

USP <1790> also acknowledges the potential use of artificial intelligence in section *6.3 Automated Visual Inspection*, citing an Nanonets article from Khan J. "*Everything you need to know about Visual Inspection with AI*" and a presentation from PDA Visual Inspection Forum 2019 from Delgado J. "*Deep Learning Development and Qualification in Automated Vision Inspection Technology for Parenteral Pharmaceutical Drug Products.*" (17):

*"Significant effort is required to program these systems and to test their performance against a range of known defects, as well as acceptable containers. Recent experimentation with artificial intelligence (AI) or deep learning (DL) has shown promise in improving the discrimination ability of the AVI system (64). Early studies show both reduced false reject rates and improved detection performance (65). The use of these tools may also reduce the time to develop a recipe for a new product or container."*

Notably, EP chapter *5.21 Chemometric methods applied to analytical data* provides a whole chapter with recommendations on good chemometric practices. While the intent of this chapter is not directly

applicable to visual inspection, it offers valuable considerations on data, maintenance of models, assessment and validation of models and possible uses and critical aspects of different models. A section about Artificial Neural Networks (ANN) also covers Convolutional Neural Networks (CNN), a technique relevant to image analysis tasks. In the context of chapter 5.21, possible uses for CNNs are in the scope of chemical imaging and interpretation of spectroscopic data. Nevertheless, the ANN principles and critical aspects, such as over- and underfitting of training data, are transferable to other use cases such as in visual inspection via AI/ML.

## 2.2 Visual inspection

Visual inspection is a critical in-process control (IPC) step to ensure the absence of defects in parenteral drug products. The inspection process should be designed and validated to ensure that every lot of the pharmaceutical drug product is safe for patients. Every single container of the lot has to be inspected, hence it is also referred to as "100% inspection". Nonconforming units must be rejected and removed during VI. Batch sizes vary greatly in the number of filled units depending on the product type, formulation, and manufacturing process. Typical batch sizes can range from a few filled units (ATMPs, orphan drugs) to several hundred thousand of units. In both cases, all filled units need to undergo visual inspection before labelling and packaging of the product can start.

While small to mid-size manufacturing operations mainly use MVI or semi-automatic visual inspection (SAVI), larger operations tend to rely on AVI to accommodate the need for high throughput when large batches are produced. Limitations inherent to human inspection such as operator subjectivity, constraints in throughput, tedious training and qualification of personnel, cost, and errors due to monotonous work have driven the evolution toward more automated VI approaches. In AVI, human operators are replaced by camera stations that acquire images which are then fed to software that makes a pass or fail decision based on a vision recipe that is specific for an AVI machine-configuration and product. Replacing human eyes (sensing device) with machine vision and the human brain (judgement) with software both has advantages and limitations. Creation of vision recipes is tedious, needs to be tailored to a specific AVI-equipment / product combination and does not always account for variability in packaging materials, product appearance or visual artefacts such as reflections, glaring, or air bubbles within the product that are formed due to setting the product in motion for detection (18). These factors may lead to rejection of good product by the software and may hence lead to costly reinspection by MVI or SAVI or even decrease batch yield if rejects are discarded.

### 2.2.1 Inspection methods and technologies

As just described, there are three visual inspection methodologies that are commonly used, depending on the scale and complexity of the manufacturing process and the product to be inspected:

1. Manual Visual Inspection (MVI): This is the reference method and relies on trained and qualified human operators to identify defects. Containers are manually set in motion to aid detection of stationary particles. Although MVI can be highly effective, especially with difficult to inspect products, human subjectivity, low throughput and inspector fatigue remain challenging.
2. Semi-Automated Visual Inspection (SAVI): Material handling to the operator is automated, the inspection and decision making is still performed by human operators. These systems enable control of the presentation of units to the operator and can offer magnification or different light setups such as Tyndall lighting.
3. Automated Visual Inspection (AVI): AVI uses automated material handling with camera stations and digital image processing to automatically accept or reject the inspected units. While AVI increases throughput, initial capital investment and time needed to configure the vision recipe

for specific products is often tedious and might not fully account for normal variations in product appearance or visual artifacts (reflections or air bubbles due to motion…).

USP chapter <1790> provides further details on MVI, SAVI and AVI. In recent years, a new equipment type in the form of a "Vision Robot" as alternative for smaller batch sizes, second stage inspection or handling of difficult to inspect products such as large infusion containers appeared. A robot arm combined with computer vision and deep learning mimics the motion of human inspectors to thoroughly inspect the unit from different angles and lighting conditions (19–22). **TABLE 3** shows the evolution of inspection equipment.

| Evolution of Visual Inspection Equipment | | | |
|---|---|---|---|
| MVI[a] | SAVI[b] | AVI[c] | RVI[d] |
|  |  |  |  |

*Table 3 Evolution of Visual Inspection equipment: From Manual- (MVI), to Semi-automated, (SAVI), to Automed- (AVI), to Robotic Visual Inspection (RVI). Illustrative equipment images: a) Adelphi Apollo II Liquid Viewer (23), b) Körber V90+ (24), c) Brevetti K15 DR (25), and d) Stevanato Group VRU (26).*

### 2.2.2 Difficult to Inspect Products (DIP)

Some parenteral drug products are more difficult to inspect due to their dosage form or primary packaging container characteristics and are referred to as "Difficult to Inspect Product" or "DIP". Broadly speaking, DIPs do not fall into the product category of clear aqueous solutions in a transparent container that can be easily inspected. Increased preventive measures by implementing a holistic contamination control strategy (CCS) (13), using statistical sampling followed by destructive testing, or the use of other imaging methods such as x-ray detection (27,28) are necessary for DIP inspection. Common industry approaches to inspection of DIP are listed in **TABLE 4**.

| Category | USP 790 (with) modification | Destructive method number* | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **DIP - Formulation type** | | | | | | | | |
| Deeply colored solutions (opaque) | | | x | | x | | | |
| Emulsion | | | | x | | x | | |
| Gels | x | | | | x | | | |
| Lyophilized | | x | x | | x | | | |
| Powders, API | | x | x | | | | | |
| Suspension (non-settling / opalescent suspension) | x | | x | x | | | | |
| Suspension (settling or very milky suspension) | | | | x | | x | x | |
| Viscous solutions | x | | x | | | | | |
| **DIP - Container type** | | | | | | | | |
| Light protective glass | x | | x | | x | | | |
| Large volume fills in glass | x | | x | | | | | |
| Plastic flexible bags | x | | x | | | | | |
| Blow-Fill-Seal | x | | x | | | | | |
| Plastic syringes | | | x | | x | | | |
| Small dosage forms (<1 ml) | x | | x | | x | | | |
| **DIP – Unique Product Type** | | | | | | | | |
| Cartridges | x | | | | | | | |
| Drug substance-eluting stents | | | | | | | | x |
| Multiple chamber syringes | x | | x | | | | | |
| Empty sterile containers | x | | | | | | | x |
| Implants – solids for extended release | | | | x | | | | |
| Implanted devices | | | | | | | | x |
| Injection pens | x | | x | | | | | |
| Medical devices and pump systems | | | | | | | | x |
| Micro-needle arrays | | | | | | | | x |

*Table 4 Common approaches to inspection of DIPs, table based on PDA Technical Report 79 - Particulate Matter Control in Difficult to Inspect Parenterals (29). The table is grouped into three main categories (DIP formulation-, container-, and unique product type) and assigned with commonly used VI methods. * Method number: 1 = Reconstitution, 2 = Filtration, 3 = Clarification, 4 = Transfer/Dilution, 5 = Sieve/Mesh, 6 = Panning and 7 = Rinse/Flush and filtration.*

Inherent protein particles of biologics, for example aggregation of monoclonal antibodies that can be induced by agitation, may be acceptable when their presence is expected and are part of the clinical profile (30).

### 2.2.3 Probability of Detection (PoD)

The detection of defects is probabilistic in nature, irrespective of the inspection method (human or camera). This means that even if 100% of containers are inspected, not all defects can be detected during visual inspection. As mentioned above, pharmacopoeias account for this fact in the wording of the acceptance criteria ("…essentially free…", "…practically free…", "…free from readily detectable…"). The detection probability for a specific defect will depend on visual attributes (size, shape, color, density, reflectivity,…) and optical characteristics of the surrounding product and packaging (30) as well as the inspection methodology (31), environment (minimizing distraction during operation) and human factors (experience, stress, fatigue, repetitiveness of the task, physical and mental health) (32).

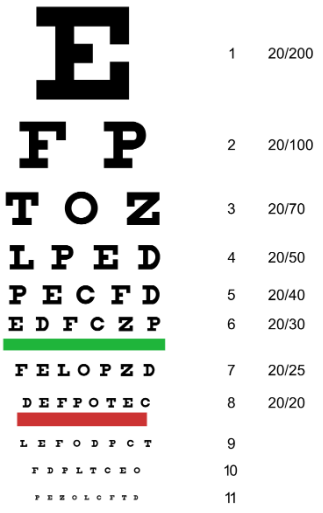


*Figure 1 Snellen chart with different sized letters used in testing of visual acuity. Source: Wikipedia*

Visual acuity is the ability to recognize small details with precision, with a "20/20 vision" describing someone's ability to recognize small details at a far distance, for example by identifying different sized letters on a Snellen chart (33), see **FIGURE 1**. A human with 20/20 vision is technically able to identify a gap of 50 µm at a distance of around 18 cm (34). However, the PoD of a 50 µm particle within a clear solution in a 10 ml vial inspected via the pharmacopoeia MVI method is around 4% (35). This low PoD illustrates that being able to precisely detect small differences on a Snellen chart (2D space) for a short period is significantly easier than detecting moving objects in a 3d space through media with different optical density for an extended time.

Operators performing VI need to undergo qualification (13). To qualify operators for MVI or to compare the performance of AVI to MVI, qualified test kits made up of containers without (80-90 %) and different defect types (10-20 %) are used. A threshold study, usually using the Knapp-methodology, is used to define the PoD for different types and sizes of particles. The Knapp-methodology recognizes that not all particles can be detected reliably. To determine the probability of detection, every unit of a test set is inspected multiple times by a panel of qualified inspectors to obtain statistical confidence. Based on the PoD of the unit, the units are then sorted into one of three zones: Reject Zone (RZ) with a PoD > 70 %, Gray Zone (GZ) with a PoD between 30-70 %, and Accept Zone (AZ) with a PoD < 30 %, see **FIGURE 2**. Units in the RZ are readily rejected by a qualified operator at least 70 % of all cases, GZ units are below the threshold of reliable detection and accepted as good, and AZ units that are accepted as good. Knapp also defined a Reject Zone Efficiency (RZE), which is a practical measure of inspection performance that incorporates the known uncertainty of VI and is the average reject rate for defective units in the RZ (units rejected in the RZ / number of units in RZ). When comparing alternative VI methods, such as AVI with an AI/ML subsystem, compared to MVI, then the RZE of the alternative method needs to be equal or better than the RZE of MVI to be suitable.

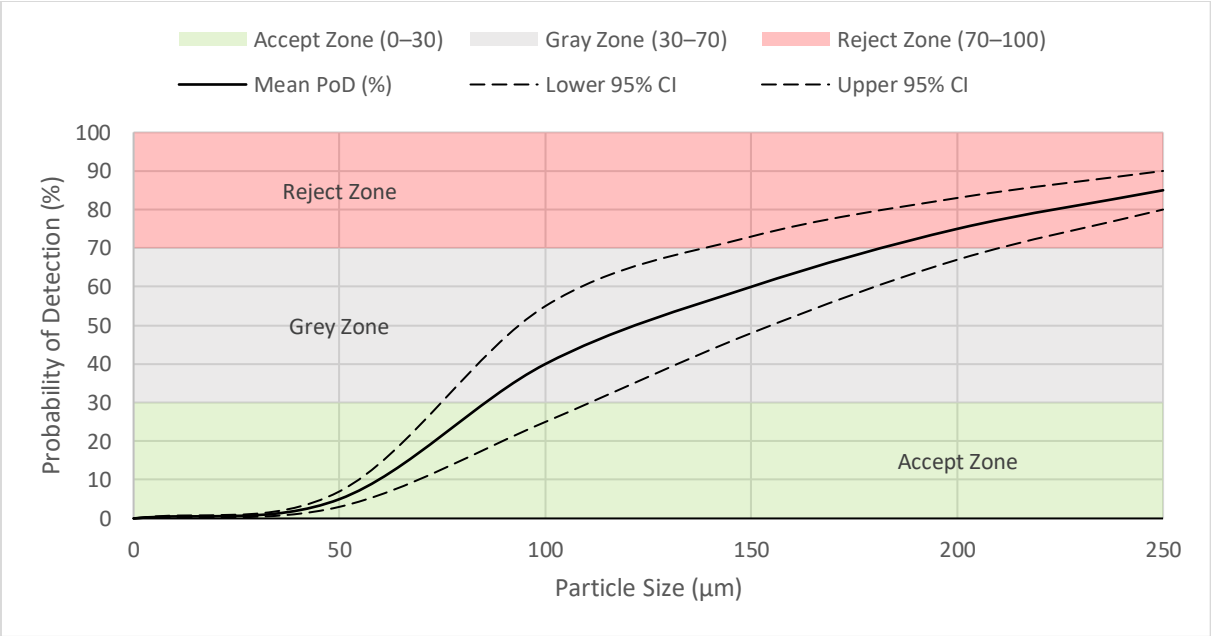


*Figure 2 Example of a curve illustrating the relationship between specific defect sizes, for example a particle, and the corresponding Probability of Detection (PoD) based on (16,36,37). The graph defines three zones: Accept Zone (0-30 % PoD), Grey Zone (30-70 % PoD) and Reject Zone (70-100 % PoD), along with corresponding upper and lower 95 % confidence intervals.*

There is no universally agreed upon definition of what "visible" exactly is. However, USP chapter <1787> states that *particles larger than 100 µm are generally considered as visible particles*. Particles in range of 2-100 µm are categorized as "sub-visible". Particles that cannot be seen reliably are categorized as "Visible Grey Zone" and roughly fall into the size range of 50 to 150 µm, depending on the type of particle, product, container and so on – hence the Grey Zone is better defined by the PoD rather than size. The RZ is essentially used to define what is "visible" since no specific size definition (apart from the general 100 µm statement in USP <1787>, is provided in pharmacopoeias. There is no specific size threshold, because the detectability can change depending on the particle, product, or container types. If a particle has a PoD greater than 70 %, then it is defined as visible.

### 2.2.4 Visual inspection process workflow

Visual inspection starts with 100% inspection of all filled containers during production, at a point where defects are most easily detected – prior to labelling and packaging. This step can either be performed in-line, at-line, or off-line by any of the inspection methods described before. A typical single-stage VI process as described in USP chapter <1790> is illustrated in **FIGURE 3**.

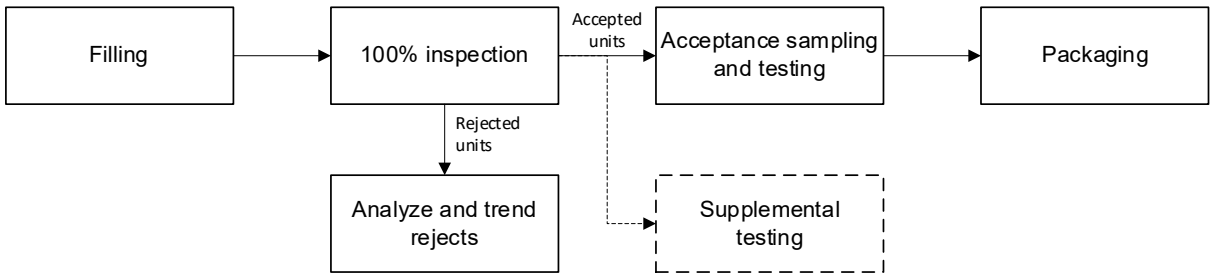


*Figure 3 Typical single-stage VI process flow chart according to USP <1790>. The process begins with filling of containers, followed by a 100% inspection (non-destructive) step where units are either accepted or rejected. Accepted units undergo acceptance sampling and testing before continuing to labeling and packaging. Rejected units are subjected to further analysis and trending. Supplemental destructive testing is necessary when product or container characteristics hinder visual inspection (DIPs such as opaque containers, lyo-cakes, powders,…).*

After 100% inspection, accepted units are subject to acceptance sampling and testing by the quality control unit to confirm the results of the initial 100% inspection. This statistical sampling is most commonly conducted according to ANSI/AQ Z1.4 or ISO 2859-1 (5). AQL testing is usually performed via MVI rather than SAVI or AVI (5) because humans are able to more readily detect unexpected, rare or new defects compared to (semi-)automated systems. In certain cases, tightened AQL plans may be applied, for example in the case of atypical results.

Limits for rejection rates (% defective units in a batch) must also be established for every defect category (for example critical, major, minor) to demonstrate process control and to identify atypical lots. It may even be appropriate to define rejection rate limits for individual critical defects (38). Functional defects of the container-closure system pose a higher risk to product quality and patient safety (possible microbial ingress leading to loss of sterility), whereas cosmetic defects are appearance flaws that do not affect product functionality but may fail to meet the high aesthetic standards of certain markets ("Japan Quality") (39). Examples of defect classes and typical action limits for both 100% inspection and AQL testing, based on the ECA Good Practice Guide "Visual Inspection of Medicinal Products for Parenteral Use" (38), are provided in **TABLE 5**.

| Defect class | Description | Typical Action limits for 100% inspection | Typical Action Limits for AQL testing |
|---|---|---|---|
| Critical defects | May cause a lack of sterility, container integrity or cause harm to patients | 0,5 to 1,0 % | 0,01 to 0,10 % |
| Major defects | May alter the content or the function of the product | 1 to 3 % | 0,10 to 0,65 % |
| Minor defects | Defects that do not affect patient health or product functionality | 3 to 5 % | 1,0 to 4,0 % |
| Cosmetic defects | Appearance defects for products intended to be placed to culturally sensitive markets | n.a. | n.a. |

*Table 5 Example of defect classes and typical action limits and AQL limits based on the ECA Good Practice Guide "Visual Inspection of medicinal products for parenteral use" (38).*

USP <1790> further addresses remediation and alternative practices for repeating the 100% inspection, followed by acceptance sampling. In particular, a two-stage inspection may be considered "*in cases where an assignable cause, such as the formation of air bubbles or specific container/closure variation, results in a high false-rejection rate*." Under this multi-stage inspection approach, ejected units undergo a confirmation inspection (Stage 2) prior to the final inspection outcome, as depicted in Figure 4.

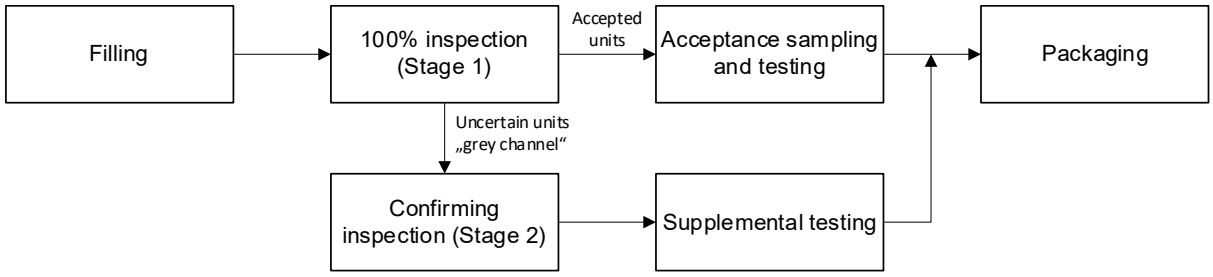


*Figure 4 Two-stage VI process according to USP <1790> introducing a confirming inspection (Stage 2) for units of uncertain disposition. These units undergo a secondary inspection, for example extended MVI, to increase sensitivity and to reliably confirm acceptance or rejection of the unit.*

### 2.2.5 Visual inspection program

A robust Visual Inspection process workflow can be integrated into a broader contamination control strategies, such as the EU Annex 1 Contamination Control Strategy (CCS) (13) or the FDA's Visual Inspection Program (40). Data generated through VI is analyzed and interpreted to help prevent contamination in upstream manufacturing processes. Since particulate matter may act as a substrate for microbial cross-contamination, the risks of microbial and endotoxin/pyrogen contamination should be viewed in context with results from the environmental monitoring of clean rooms.

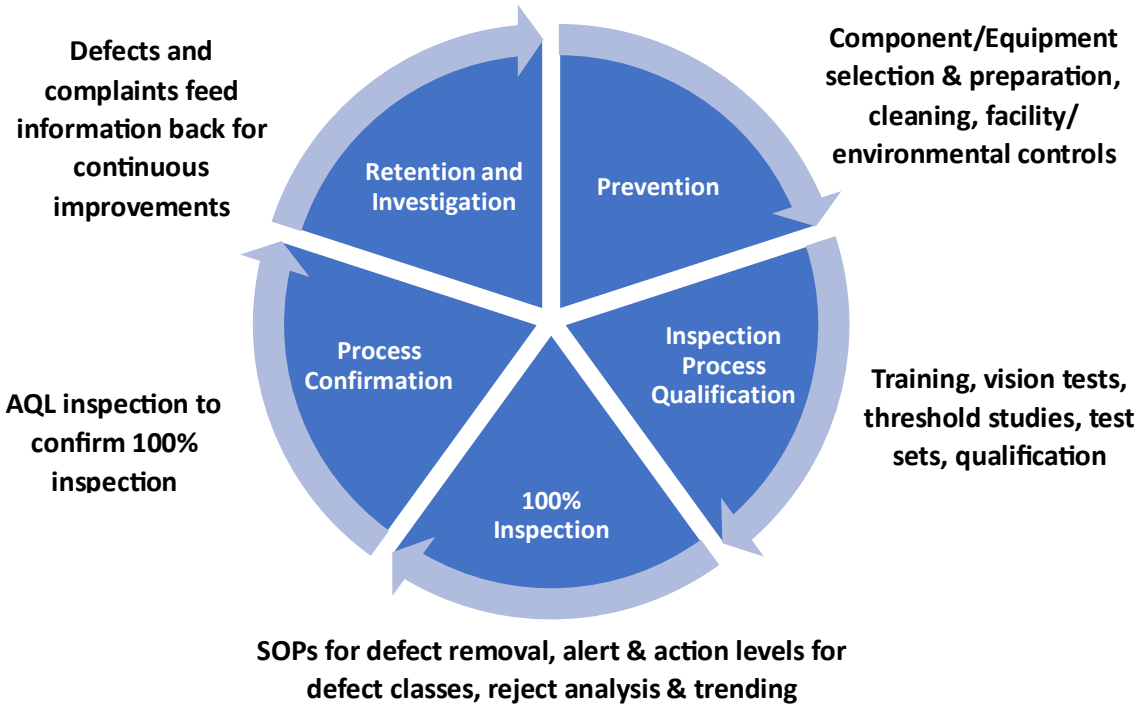


*Figure 5 A holistic lifecycle approach to visible particle control of injectable products according to FDAs draft guidance on "Inspection of Injectable Products for Visible Particulates".*

The FDA Draft Guidance on the "Inspection of Injectable Products for Visible Particulates" (40) is aligned with USP monographs <790> and <1790> and acknowledges the probabilistic nature of the VI process. Consequently, the VI process should be embedded in a comprehensive lifecycle model, see **FIGURE 5**, which encompasses all areas relevant to defect prevention, detection and control, ultimately driving continuous improvement (41–43). In this model, data from the VI program is at the core of the control strategy, with trend analyses (e.g., statistical process control [SPC] charts) and AQL-based action limits ensuring the VI process remains under control. Defect characterization helps to provide insights into upstream processes and supports the continuous improvement of the overall production process. Additionally, findings from the visual inspection process support updates to the master defect library, aid in the preparation of defect kits, and contribute to both operator and machine qualification.

## 2.3 Artificial Intelligence

Artificial Intelligence (AI) has evolved into a scientific discipline that accelerates many modern technological advancements. It is a paradigm shift from traditional rule-based software development approaches, offering models that can learn from data, identify patterns, and make predictions or decisions with a degree of autonomy. This evolution is particularly relevant in the context of the USP <1790>, which acknowledges the potential of AI to enhance AVI-systems. As highlighted in the monograph, AI may not only offer improvements in inspection throughput and consistency but also show promise in reducing false rejection rates by better handling normal variations in product containers.

AI is broadly recognized as the ability of machines to exhibit intelligence through learning, problem solving, and decision making. AI, ML, and Deep Learning (DL) encompass various techniques that collectively fall under the broader scope of science. As displayed in **FIGURE 7**, DL is a subfield of ML and ML in turn is a subfield of AI. The common goal of ML/DL is to extract meaningful information from data through patterns (44). Data is anything that can be represented numerically, for example: prices, text, sound recordings, images or videos (44). The latter examples are especially relevant to image processing in AVI where patterns of different types of defects need to be recognized. Traditionally, images from AVI machines are analyzed by software that was created as an "expert system" that captures the thinking and rules of vision experts, engineers and scientists with a set of discrete rules. These systems are difficult to build and maintain, since the process of producing the rules ("feature engineering") requires significant human intervention, time, effort and expertise. ML addresses this problem by finding the underlying rules algorithmically (44,45). Deep Learning represents a class of algorithms within ML that use layered, or "deep", neural networks to extract features from raw data. This emerging field can be dated back to the 1940s and has undergone several different brandings of the name such as "Cybernetics" in the 1940s-1960s, "Connectionism" 1980s-1990s, "Artificial Neural Networks" and "Deep Learning" since the 2000s (46). "Deep" refers to the fact that the learning algorithm uses multiple layers of neurons that appear deep when drawn stacked up vertically (44,47), see **FIGURE 6**. These interconnected "neurons" or nodes are organized in layers and information is processed with different weights applied to through these layers in the form of a "Artificial Neural Network" (ANN) (48), see Figure 6. The most common variants of neural networks (NN) are:

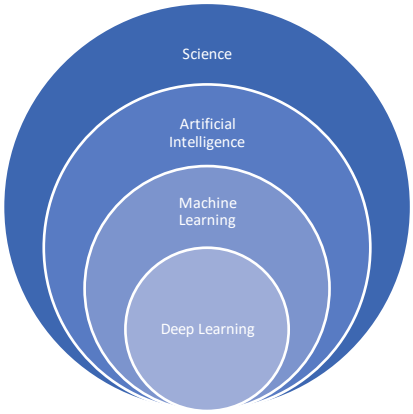


*Figure 7 Hierarchical relationship between science, AI, ML, and DL.*

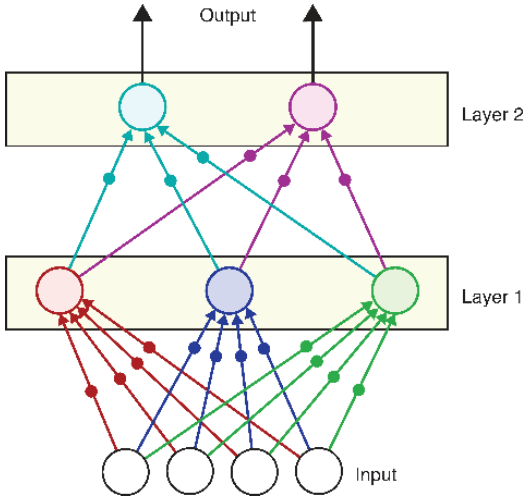


*Figure 6 A "deep neural network" with four input values that go upwards into a layer of "artificial neurons" that are fed to another layer that generates an output. The connections between the neurons are weighted, meaning that the value is multiplied before it reaches the next neuron. Figure from (35).*

- Convolutional neural networks (CNN) which are optimized for processing of grid-like data as found in images (49), and
- Recurrent neural networks (RNN) that are suited for sequential data such as text, speech and time series where order of elements is important (50).

Given the absence of AI/ML definitions, both the EMA and FDA have published glossaries (51,52) containing their respective definitions. Key terms are put in contrast to each other in **TABLE 6**.

| Term | Definition by EMA | Definition by FDA |
|---|---|---|
| AI | Artificial intelligence. OECD definition - an AI system is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and 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. | 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. |

| Term | Definition by EMA | Definition by FDA |
|---|---|---|
| ML | 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. | A mathematical construct that generates an inference or prediction for input data. This model is the result of an ML algorithm learning from data. Models are trained by algorithms, which are step-by-step procedures used to process data and derive results. AI systems (e.g., AI-enabled medical devices) employ one or more models to achieve their intended purpose. |
| DL | Approach to creating rich hierarchical representations through the training of neural networks with many hidden layers | A specialized branch of ML that involves training neural networks with multiple intermediary (hidden) layers that operate between an input layer that receives data and an output layer that presents the final network output. Each layer learns to transform its input data into a slightly more abstract and composite representation and produces an output that serves as an input for the next layer. As data propagates through successive layers, these models are able to learn hierarchical feature representations from the input data. |
| NN | Network of one or more layers of neurons connected by weighted links with adjustable weights, which takes input data and produces an output | A computational model inspired by the structure of the human brain. It is composed of interconnected nodes, or "neurons" organized into layers: an input layer that receives data, one or more hidden layers that process and identify patterns in the data, and an output layer that presents the final network output. |
| CNN | - | A specialized deep neural network architecture that consists of one or more convolution layers that is suited for processing grid-like data, such as images. In a convolution layer, a "filter" (window or template) slides over regions of the input image to identify low-level patterns (e.g., edges) by applying convolution (a mathematical dot operation applied to the input data). Different filters can be applied to extract different features, such as edges, textures, or curves in images. Additionally, CNNs can include pooling layers, whose function is to reduce the feature dimensionality while retaining relevant features. These convolution and pooling layers get stacked on top of each other to enable this network to build up a hierarchical understanding of patterns and makes CNNs effective at tasks like image recognition and computer vision. An important aspect of this network is its ability to conserve spatial information of the original input while still performing the feature extraction. |

*Table 6 Comparison of EMA and FDA definitions of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Neural Networks (NN), and Convolutional Neural Network (CNN). Source: (51,52)*

### 2.3.1 Fundamental paradigms in Machine Learning

Core paradigms in ML are traditionally categorized into three main training approaches, depending on the input or feedback available to the algorithm:

1. Supervised Learning: Uses data (input X) that is labeled (output Y) for training with the goal to learn a general rule that maps the inputs to the desired outputs. For example, labeled images from visual inspection with all known defects as well as good units are used to train a binary classification model (accept, reject) that can be used in AVI (18).
2. Unsupervised Learning: Uses only input data (input X) without labels with the goal to discover inherent structure or hidden patterns in the data through clustering. For example, exploring images for defects that were previously not identified (18).
3. Reinforcement Learning: A model (or agent) interacts with an environment and learns to make decisions through getting rewards or penalties. For example, finding a CNN architecture through the best combination of parameters (53,54).

Combinations of these core paradigms such as Semi-supervised Learning (mixed labeled and unlabeled training data), ensemble methods (a combination of multiple base models) or the use of pre-trained models is also possible. For example, pre-trained models that learned low-level features such as detection of edges from prior training on large datasets can be fine-tuned to a specific use case such

as detection of cracks in glass vials. Fine-tuning involves updating the weights of the model using new training images (55).

Understanding these foundational concepts is essential, especially when later interpreting specialized applications such as those used in pharmaceutical visual inspection. **FIGURE 8** shows a hierarchical overview of major ML paradigms, highlighting the DL class as this is of high importance for applications in VI of pharmaceutical drug products, including examples of algorithms

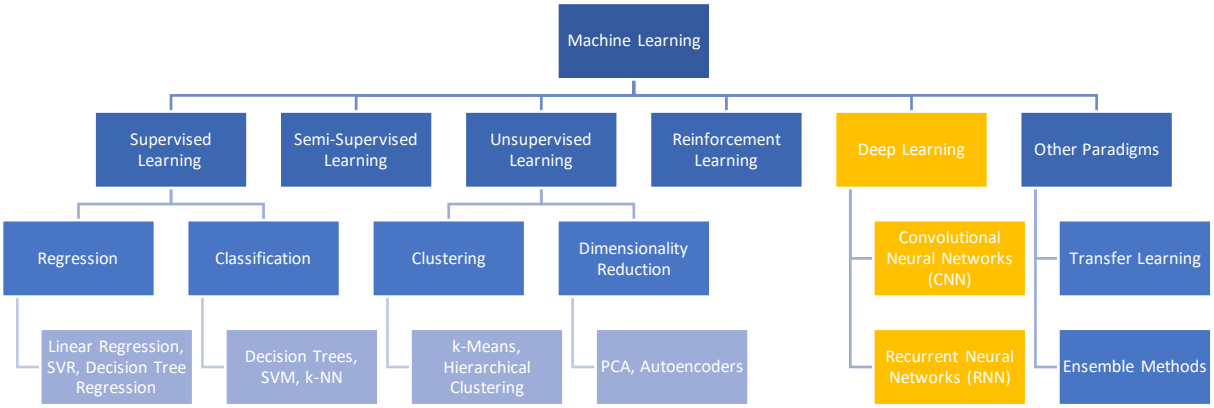


*Figure 8 A hierarchical overview of major ML paradigms, illustrating supervised, semi-supervised, unsupervised, and reinforcement learning approaches, as well as DL methods and other paradigms. Subcategories for approaches are shown, including example algorithms such as Linear Regression, k-Nearest Neighbors (k-NN), k-Means, Q-Learning, Convolutional Neural Networks (CNN), and more.*

### 2.3.2 Model development

A concise framework for ML model development can be found in the flowchart from EP monograph *5.21 Chemometric methods applied to analytical data*, which illustrates important steps of the supervised model development process, see **FIGURE 9**. While the intended use and some terminology differ from the use case for visual inspection, principles such as data preparation, data partitioning, and the iterative model-design process are common.

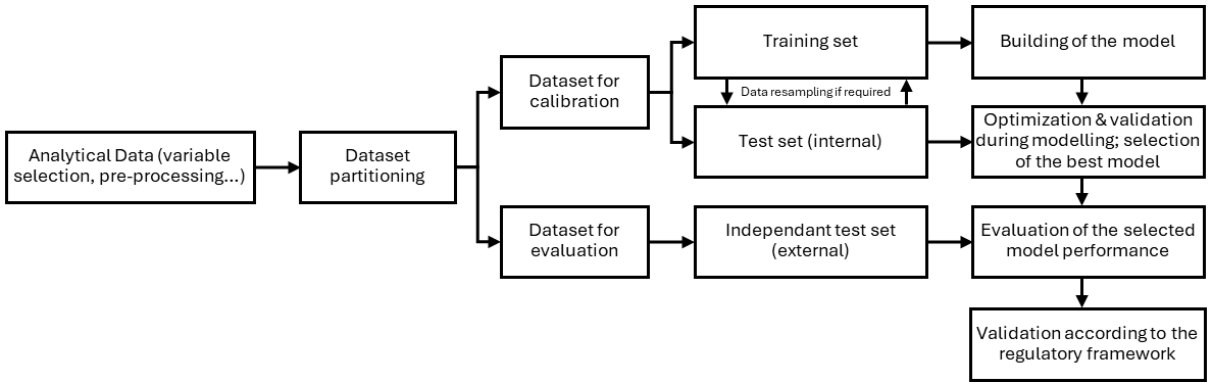


*Figure 9 Flowchart for the validation process of a supervised chemometric model according to EP (11.1) monograph 5.21 Chemometric methods applied to analytical data.*

To develop a candidate model and estimate its performance, sufficient data is needed for training. For supervised learning, data collection is followed by labeling of individual data points, data cleaning and pre-processing to establish a balanced dataset. Data needs to be managed with the same care as the

code itself since the learning data is part of the model, hence additional controls for data acquisition, selection, classification, cleansing, and augmentation are necessary; see **4.2.2 DATA GOVERNANCE**.

A properly labeled dataset (~calibration samples) provides the "ground truth" that the ML model uses to check its predictions for accuracy and to continue refining its algorithm. Errors in data labeling impair the quality of the dataset and any resulting predictive model training it is used for. In its simplest form, the image is classified by assigning a label to an image. Advanced labelling tasks such as drawing a bounding box (BBox) or by using pixel-wise segmentation to locate an object with its boundaries are available as well but requires specialized annotation tools and trained personnel.

If the amount of data needed for training is not sufficient or if statistical distribution of certain populations is not representative, techniques like data augmentation can be used to create a balanced dataset. Data augmentation techniques include basic image manipulation (geometric transformations, color space transformation,…) or synthetic data creation via software (56,57).

Once the dataset is complete, its data is split into a training-, validation- and test set; see **FIGURE 10**. **TABLE 7** gives an overview of the purpose, description and usage timing of the individual datasets. The ratio for splitting depends on the amount of data available and the individual use case. It should be noted that "Validation set" is data science terminology that is commonly used by developers to monitor the model's performance to adjust hyperparameters (learning rate, number of layers, batch size, epochs, dropout rate…) and model architecture to reduce overfitting and optimize for generalization. The term "*Validation* set" might be misleading in a GMP setting and hence FDA uses the term "Tuning data" instead, EMA sticks to the computer science terminology "Validation dataset". "Training data" and "Validation data" are used to develop a model, "test data" is used to evaluate the performance of the trained model, see **FIGURE 10**.

| Data Type | Description | Purpose | Usage Timing |
|---|---|---|---|
| Training dataset | Data used to teach the model to learn patterns, relationships, and parameters from the input features | Learn model parameters and underlying patterns | During model training |
| Validation (Tuning) dataset | Data used as a checkpoint to tune hyperparameters, to prevent overfitting, and to adjust learning strategies and decide when to stop model training | Hyperparameter tuning, model selection, and overfitting prevention | During model development |
| Testing dataset | After the model is fully trained and tuned, the test data—kept separate and unseen during training—is used to provide an unbiased evaluation of the model's performance. It serves as a proxy for how the model will perform on entirely new, real-world data | Final performance evaluation and generalization assessment | After model training and tuning |

*Table 7 Overview of data subsets in machine learning, highlighting their primary purpose and the phase during which each subset is used.*

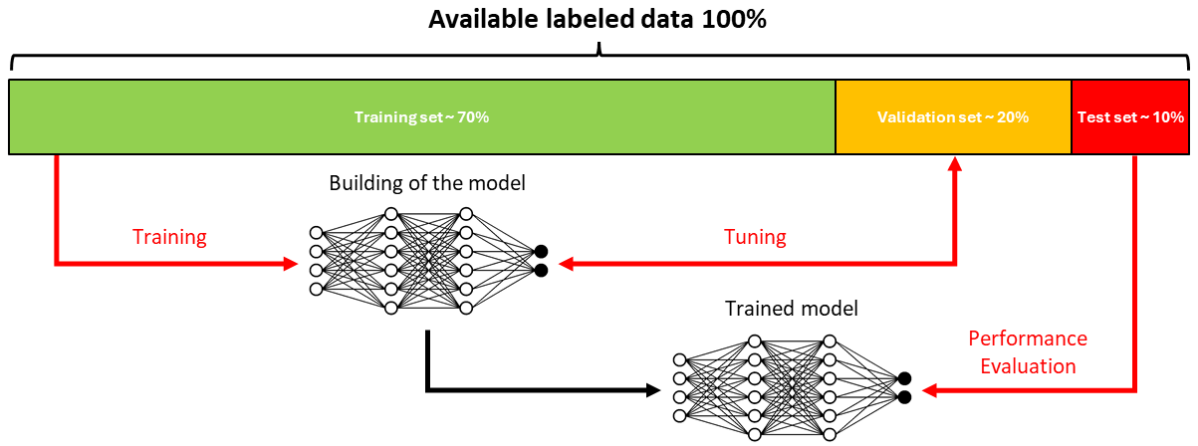


*Figure 10 Schematic representation of a common data splitting strategy in supervised ML, showing 70% of the labeled data allocated to training the neural network, 20% reserved for tuning of hyperparameters, and the remaining 10% used for final performance evaluation of the fully trained model.*

### 2.3.3 Model performance evaluation

Training a model with data is not enough, it needs to be evaluated to understand whether it needs to be improved or whether it performs well in the real world or not. Performance metrics are used to evaluate the predictive accuracy and reliability of ML models. For classification problems, for example to categorize defects in VI, a confusion matrix is used to show the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). This type of table offers a direct view of the performance and highlights if the model tends to misclassify between specific classes. For example, the confusion matrix (= error matrix) can reveal systematic errors where one class is consistently mistaken for another, see **TABLE 8**.

| | Predicted: No defect | Predicted: Crack | Predicted: Chip | Predicted: Scratch | Predicted: Particle |
|---|---|---|---|---|---|
| **Actual: No defect** | 100 | 0 | 0 | 0 | 0 |
| **Actual: Crack** | 0 | 65 | 0 | 35 | 0 |
| **Actual: Chip** | 0 | 0 | 100 | 0 | 0 |
| **Actual: Scratch** | 0 | 0 | 0 | 100 | 0 |
| **Actual: Particle** | 0 | 0 | 0 | 0 | 100 |

*Table 8 Example multi-class confusion matrix for a defect classification ML model. The model's predictions (columns) and the ground truth (rows) for different defects are tabulated with diagonal entries presenting correct classification, while off-diagonal cells show misclassification. The model misclassified 35 cracked instances and classified them wrong as "scratch".*

Several key metrics can be derived from the confusion matrix. Collectively, these metrics help address challenges posed by biased class distributions. **TABLE 9** and **TABLE 10** summarize key performance metrics used in the evaluation of ML models, detailing their calculation methods, interpretive value, and associated limitations.

| | Predicted Positive | Predicted Negative | |
|---|---|---|---|
| **Actual Positive** | True Positive (TP) | False Negative (FN)<br>*Type 2 error = Patient risk* | **Sensitivity (= Recall = True Positive Rate)**<br>$\frac{TP}{(TP+FN)}$ |
| **Actual Negative** | False Positive (FP)<br>*Type 1 error = Business risk* | True Negative (TN) | **Specificity (= True Negative Rate)**<br>$\frac{TN}{(TN+FP)}$ |
| | **Precision (= Posit. Predictive Value)**<br>$\frac{TP}{(TP+FP)}$ | **Negative Predictive Value**<br>$\frac{TN}{(TN+FN)}$ | **Accuracy**<br>$\frac{TP+TN}{(TP+TN+FP+FN)}$ |
| | **False Positive Rate (FPR)**<br>$\frac{FP}{(TN+FP)}$ | **False Negative Rate (FNR)**<br>$\frac{FN}{(TP+FN)}$ | **F1-Score**<br>$\frac{2\,x\,(Precision\,x\,Recall)}{(Precision+Recall)}$ |

*Table 9 Illustration of a binary classification confusion matrix with True Positive (TP), False Negative (FN) or Type 2 errors, False Positive (FP) or Type 1 errors, and True Negative (TN) results. It also shows how common metrics like sensitivity, specificity, precision, negative predictive value, and accuracy are derived from these values (58,59).*

| Metric | Formula | Interpretation | Limitation |
|---|---|---|---|
| **Accuracy** | $\frac{TP+TN}{(TP+TN+FP+FN)}$ | Proportion of correct predictions among all samples. | Can be misleading in imbalanced datasets (e.g., predicting all samples as the majority class may still yield high accuracy). |
| **Precision** | $\frac{TP}{(TP+FP)}$ | Proportion of predicted positives that are actually positive. | Ignores false negatives; a model can achieve high precision by predicting fewer positives but missing many actual positive cases. |
| **Sensitivity (Recall)** | $\frac{TP}{(TP+FN)}$ | Proportion of actual positives correctly identified.<br>( = True Positive Rate = TPR ) | Ignores false positives; predicting all samples as positive can yield high recall but may produce many false alarms. |
| **Specificity** | $\frac{TN}{(TN+FP)}$ | Proportion of actual negatives correctly identified. | Ignores false negatives; a model might miss many positives yet maintain high specificity if it correctly identifies most negatives. |
| **F1-Score** | $\frac{2\,x\,(Precision\,x\,Recall)}{(Precision+Recall)}$ | Harmonic mean of precision and recall, balancing both metrics. | Does not account for true negatives; can be less informative when correctly identifying negatives is also critical (e.g., heavily imbalanced data). |

*Table 10 Common performance metrics derived from the confusion matrix, their formulas, interpretations, and limitations.*

Error analysis is an essential component of model evaluation, as it provides a detailed examination of misclassifications. In binary classification, misclassifications fall into two categories: false negatives (FN)

and false positives (FP). Understanding the distribution and causes of these errors is critical for refining overall model performance. Refer to section **4.1.1** for information on inspection errors.

Bias refers to the simplifying assumptions a model makes about the underlying data, which often misses relevant relationships (underfitting). Variance, on the other hand, measures how sensitive a model is to fluctuations in the training data. Too much variance causes the model to fit even the random noise in the dataset, resulting in poor generalization (overfitting). Underfitting is when the model's simplicity prevents it from finding important patterns between input and output data, and overfitting is when the model's complexity leads it to learn noise or just remember the training data instead of the overall pattern. **FIGURE 11** shows the bias-variance tradeoff and the relationship to a model's complexity (60).

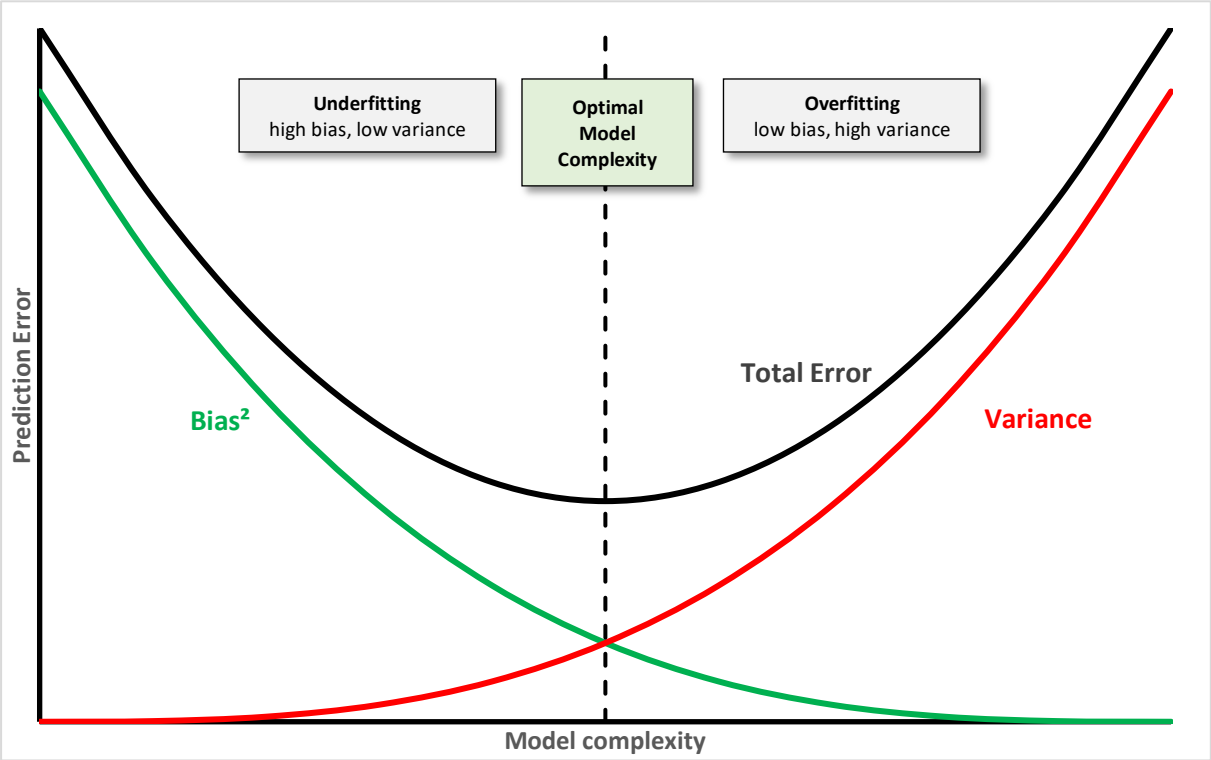


*Figure 11 Illustration of the bias-variance trade-off: as model complexity increases, bias (green curve) decreases while variance (red curve) grows, resulting in an optimal point (vertical dashed line) where total error (black curve) is minimized.*

## 3 Methods

This master thesis addresses four specific research questions (RQ1-4) listed in **TABLE 11**.

| ID | Research Question | Motivation |
|---|---|---|
| RQ1 | What are potential use cases of AI/ML software in visual inspection of parenteral drug products? | To understand how limitations of widely used visual inspection methods can be addressed with AI/ML. |
| RQ2 | What challenges and opportunities are associated with the adoption of AI/ML in GMP-regulated environments? | To find aspects of AI/ML that need to be considered in a GxP regulated area. |
| RQ3 | Does the current legal framework in Europe and the United States of America meet the need of the pharmaceutical industry for AI/ML guidance in GMP? | To identify how this new technology is viewed from the perspective of a regulatory body as well as from industry. |
| RQ4 | What AI/ML best practices can be derived from the medical device industry? | To evaluate if AI/ML best practices from the medical device industry can be adopted in GMP environments. |

*Table 11 Research questions and motivation for the use of AI/ML in pharmaceutical manufacturing.*

The literature used in this work was collected from various sources to have a holistic understanding of the use of AI/ML in visual inspection of parenteral drug products. A large portion of the available

information on AI/ML use in pharmaceutical manufacturing is not published in peer-reviewed articles and hence not available in conventional academic journals or not accessible without considerable expense. Due to the fact that AI/ML use-cases gained popularity recently, there is currently no sector-specific framework in place to regulate the use of AI/ML in pharmaceutical manufacturing. Apart from utilizing the available peer-reviewed papers, grey literature was a significant component in answering the research question above. Organizations such as the International Society for Pharmaceutical Engineering (ISPE), the Parenteral Drug Association (PDA), and the European Compliance Academy (ECA) provide expert knowledge in this area. These organizations are recognized by the Pharmaceutical Inspection Co-operation Scheme (PIC/S), which plays a critical role in harmonizing Good Manufacturing Practices (GMP) globally, making it a key reference point for regulatory compliance in pharmaceutical manufacturing. Collaborations between PIC/S and organizations like ISPE, PDA, and ECA lend significant credibility and relevance to the grey literature utilized in this work. This approach ensures that the literature review captures a comprehensive view of the field, integrating both the technological and regulatory aspects that are crucial for the application of AI in the visual inspection of parenterals in the pharmaceutical industry. Furthermore, these organizations play an important role in providing feedback to regulatory bodies such as EMA and FDA on draft guidelines for the use of artificial intelligence. In addition, two databases and a search engine were queried for literature: PubMed, IEEE Xplore and the Google Scholar. Keywords and phrases were used as search strings, using logical operators for advanced search, and formatted to the need of the search fields of the individual database: *((visual inspection) AND (artificial intelligence OR machine learning) AND (manufacturing OR drug OR parenteral))*.

Questions related to the evolving regulatory frameworks were answered by reviewing AI/ML related documents from official homepages of the European Medicines Agency (EMA) and the United States Food and Drug Administration (FDA). Furthermore, videos of public presentations and workshops by these agencies were reviewed for further information on the evolving regulation. Best practices from the medical device industry were reviewed by accessing the AI/ML related guidance documents from FDA, the Danish Medicines Agency (DKMA) as well as the German Notified Bodies Alliance (IG-NB).

Zotero software was used to organize all search results, extract metadata and export formatted references.

## 4 Results and Discussion

The results of the literature review and its discussion is structured in four sections that are referring to research question RQ1-4:

- Section **4.1** focuses on use cases of AI/ML in visual inspection of parenteral drug products
- Section **4.2** explores the challenges and opportunities associated with AI/ML in a GMP setting
- Section **4.3** reviews the EMA and FDA regulatory framework around AI/ML
- Section **4.4** examines AI/ML best practices in the medical device industry

### 4.1 AI/ML in visual inspection of parenteral drug products: use cases (RQ1)

Visual inspections of products are usually conducted to make decisions regarding the suitability of a product and to gain data on the manufacturing process (61). As a type of quality inspection, VI is widely used in many different manufacturing industries such as electronics, aerospace, automotive and as part of maintenance operations to ensure quality and safety.

In pharmaceutical manufacturing, 100% visual inspection of every drug product container is a mandatory in-process control that is followed by statistical acceptance sampling testing (AQL) of the finished product before release. As a critical quality control step, VI is mandatory to ensure the absence

of defects in the product or its primary packaging. While compendial requirements mainly address the topic of visible particulate matter, Good Manufacturing Practice (GMP) such as EU GMP Annex 1 addresses inspection of defects in general that may impact product quality or safety. VI can be viewed as a control step that is part of a larger program, such as a contamination control strategy, to ensure that drug products are essentially free of defects. To demonstrate process control, a defect classification and criticality system needs to be in place in order to analyze batch performance (trending). To identify atypical lots, limits on typical rejection rates need to be established either for each defect category (critical, major, minor) or for specific defect types (particles, cracks,…) (30).

Automated visual inspection machines are used in many industrial manufacturing settings to combine automated material handling with electronic sensing of product appearance (30). The container to be inspected is illuminated and usually presented to several camera stations where one or a sequence of images are acquired. Thereafter, image processing analyzes regions of interest (ROI) via a rule-based expert system to decide if the product is acceptable or if it needs to be rejected due to detected defects. The result data is documented in an electronic system and defective containers are tracked through the machine to allow ejection of defective units. Veillon R. points out, that the difference between a typical AVI process compared to an AVI process using AI/ML is limited to the image processing step (62). Instead of using simple engineered features to evaluate parts of an image, a trained neural network uses the entire image for classification.

Hütten N. et al identified various computer vision (CV) tasks and key requirements for deep learning (DL) use cases in an extensive review of 196 scientific papers that discuss DL-based models for AVI in manufacturing and maintenance. The authors reviewed and analyzed publications that use deep learning (DL) to automate classical computer vision tasks: classification, object detection, and segmentation. Key requirements outlined by the authors for the use of DL in AVI will be discussed from an GMP-angle in section **4.2** as part of RQ2. All AVI use cases that are currently addressed by DL models are summarized under the umbrella term “quality check” and further categorized as “Damage Detection” (including the subcategory of “Crack Detection”), “Completeness Checks” and the miscellaneous category “Other” that includes use cases that cannot be directly seen through quality inspection (for example predictive maintenance). The paper presents an overview of DL-models used to solve CV problems in visual inspection via: Binary classification (for example accept/reject), multi-class classification (for example crack, fiber, particle…), localization and multi-class localization. Among the papers included in the review by Hütten N. et al, no related works in the area of VI of parenteral drug products were included. This might be due to the selected exclusion criteria by the authors and the fact that little open-access literature in the area of pharmaceutical manufacturing is published.

The following DL use-cases relevant to AVI in pharmaceutical manufacturing were identified:



### 4.1.1 Use case 1: Reduction of inspection errors and cost

During 100% visual inspection of parenteral drug products, errors such as rejecting a good product or passing of a bad product may occur and may be due to the probabilistic nature of VI or due to limitations of the information available to the inspection system. These inspection errors can be grouped into type 1 and type 2 errors (59):

- Type 1 errors (false positive = FP = false reject = α error ≙ business risk) occur when a defect-free unit is rejected although no defect is present. High false rejection rates (FRR; percentage of type 1 errors) lower the yield of the batch and add to the production cost. Excessive FRR

may also disrupt the production process, cause increased workload with re-inspection or lead to a shutdown of the production line and trigger a root cause investigation.

- Type 2 errors (false negative = FN = false accept = β error ≙ patient risk) occur when a bad product is missed and passes the inspection with a defect. Manufacturers need to minimize false accept rates (FAR, percentage of type 2 errors) to keep the risk for patients receiving a defective product low.

As pharmaceutical manufacturing is a quality-based manufacturing system (61), conservative thresholds for defect detection are prioritized to minimize false accept rates (FAR). In traditional rule-based systems, this frequently results in increased overall system sensitivity, a higher false reject rate (FRR), extensive secondary (manual) inspections, and consequently, elevated operational costs, see **FIGURE 4** and **TABLE 12**. Natural variance in product appearance, for example thickness variations in molded glass, makes it difficult for rule-based systems to account for each scenario (63).

Veillon R et al highlight specific issues in discrimination between air bubbles and particulate matter as well as between cracks and artifacts (scratches) that lead to false rejection of good products (18). Difficult-to-inspect products or critical defects with a low PoD are another challenge for AVI since rule-based systems often lack specificity for such situations: Raising the sensitivity to detect these defects would also increase the sensitivity for other minor defects, which in turn may lead to higher false rejection rates.

In 2019 Tsay C and Li Z (64) explored the application of a multi-input convolutional neural network (CNN) for the inspection of lyophilized drug products. High false rejection rates due to cosmetic defects (65) such as white haze on the inside of glass vials after lyophilization ("fogging") was addressed by the use of multiple images of the same vial taken from different angles, aiming to improve defect detection accuracy. The authors reported to achieve 85-90% classification accuracy on validation data using a CNN pre-trained on the MNIST dataset followed by training. They concluded that using transfer-learning resulted in better generalizable features compared to training models from scratch.

In 2018 Zhao M et al (66) developed an approach for detecting particles in injections by combining Faster-RCNN deep neural network followed by hierarchical and K-means clustering to detect motion trajectories. Especially the multi-frame trajectory detection by the clustering technique on the image-sequence led to an average precision improvement of 12% compared to Faster-RCNN alone. Despite these advancements, the study acknowledges the need for further optimization to improve recall rates and to distinguish between different types of foreign particles. One explanation for the increased precision might be the increased information content of multi-frame analysis compared to single-frame analysis. A presentation at the 2024 PDA Visual Inspection Forum by Goodwin A and Brandenburger W (67) pointed out that regardless of the technology used in AVI (referring to rule-based systems or deep learning), the information content (entropy) of an image needs to be optimized to keep misclassification at a low rate. Factors such as camera resolution and image sharpness, sensor bit depth, number of images and optics of the system (blur) feed into the information content available for image processing. A practical example of plunger-dome inspection is presented by the authors where a DL neural network is used to extract information from an image-sequence to reduce inspection errors.

In his thesis (68), Peter Tran addresses the issue of excessive reinspection due to high FRR of lyophilized drug products in semi-automated visual inspection by using different neural network architectures such as CNN and autoencoders. Binary classification models were used to classify images as "good" or "defect" followed by a heatmap-visualization technique (Grad-CAM) to verify that the model is making the right decision for the right reasons, i.e. the model is functioning as intended. Despite challenges such as small and imbalanced dataset, the work by Tran shows that pretrained networks tend to train

faster and have more reasonable explanations to the conclusion they make. Tran also showed that autoencoders, a type of NN used in the realm of unsupervised learning, can achieve performance comparable to the pretrained models. However, simple data augmentation techniques were found to be ineffective in improving model performance. The findings suggest that DL, particularly with pretrained models and autoencoders, holds promise for fully automating the visual inspection of lyophilized products, potentially reducing the need for MVI from 30% to as low as about 6%. Furthermore, the work also makes it evident that creating datasets is a challenging and time-consuming task. Care needs to be taken to avoid introduction of bias into the datasets that are used for model training and testing.

In his 2012 thesis "Evaluating Future Biopharmaceutical Inspection Needs, Infrastructure Capability Gaps, and Technology Development Strategies", Hollander M (69) identified technological challenges of AVI machines for future biotechnological drug product pipelines. Increased viscosity, and hence difficulty to set visible particulate matter in motion, changing presentation of materials and inherent particulate matter (protein aggregation) often leads to the use MVI due to high FRR of AVI. In his case study at Amgen Manufacturing Limited, Hollander estimated the number of full-time-equivalent (FTE) human visual inspectors at the AML Puerto Rico site needed due to AVI's technological limitations, see **TABLE 12**.

| Year | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|
| FTE | 2 | 4 | 9 | 11 | 12 | 17 | 25 | 40 | 47 | 50 |
| Cost ($) | 0,25 M | 0,42 M | 0,90 M | 1,08 M | 1,19 M | 1,73 M | 2,53 M | 3,95 M | 4,69 M | 4,96 M |

*Table 12 Annual projection of FTE's and associated cost required for MVI for non-AVI validated products based on Hollander M (69). This example drastically estimates the effect of an AVI capability gap when shifting the product portfolio towards DIPs.*

Additionally, costs for the space needed to accommodate all MVI stations and changing from a 2-shift to a 3-shift work schedule would need to be accounted for and further demonstrated the importance for technological advancements in AVI for biopharmaceutical manufacturers. Martinez E pointed out that workforce and floor space requirements need to be considered in production because of the average inspection rate of MVI. If a one shift operation uses an AVI-machine with an average inspection speed of 300 units per minute, 82 inspectors with an average of 3,67 units per minute are needed (70).

### 4.1.2 Use case 2: Acceleration of vision recipe development

A vision recipe is a meticulously designed set of parameters and algorithms that guide the AVI system in detecting and classifying defects during the inspection process. It includes configurations for image acquisition, processing algorithms, defect classification rules, and decision-making logic. The vision recipe acts as a blueprint, ensuring that consistent performance in the visual inspection process. In rule-based expert systems, vision experts are needed to manually select and extract features that are then classified. Creating a robust inspection recipe demands significant expertise and effort. As Veillon R et al point out, rule-based algorithms require a lot of time and effort to fine-tune the vision recipe to find a balance between good detection sensitivity and low FRR (18). Deep learning methods may reduce the time needed for recipe development, since the algorithm does not explicitly need to be told which features to look for but rather learn these features during training; see **FIGURE 12**.

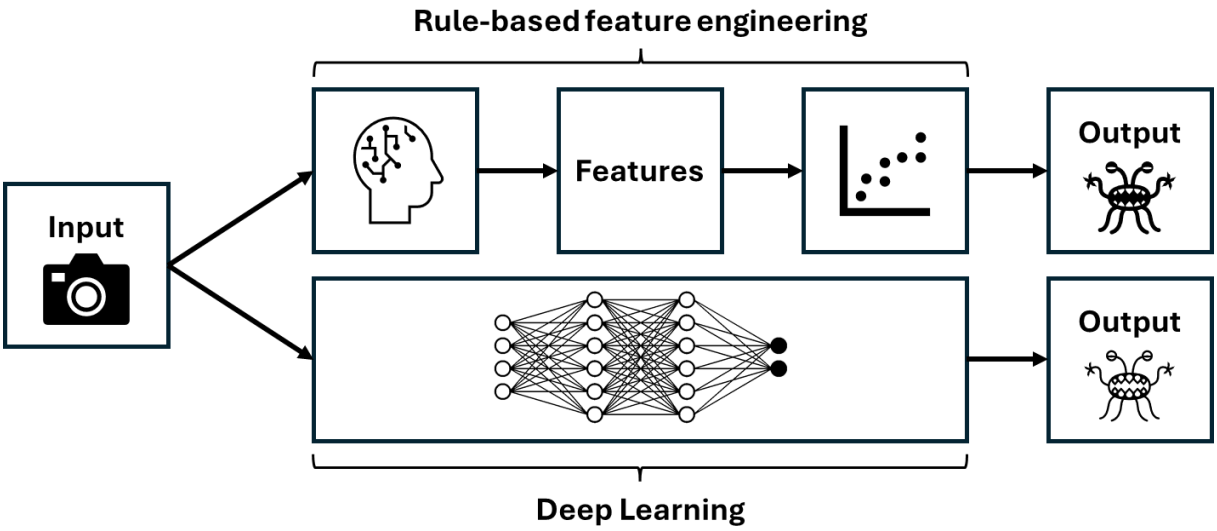


*Figure 12 Comparison of a traditional rule-based pipeline with manual feature engineering versus an DL approach.*

However, the effort needed for DL models may vary depending on how model training is performed. Transfer learning may be a good starting point compared to model training from scratch. Tran P (68), Tsay C and Li Z (64) as well as Zhao M et al (66) use transfer learning in their studies. Tsay C and Li Z described the use of transfer learning for a "base model" with a dataset such as MNIST (70 000 images of handwritten numbers in 10 categories) to learn low-level features such as edges followed by fine-tuning of the model by training on an image dataset of lyophilized drug products. Especially in situations when the available dataset is small, the approach of transfer learning may be a valuable approach to prevent issues such as overfitting of the model, yielding in poor generalization in real-world scenarios. Tran P used VGG16 and ResNet50 DNN architectures for defect detection in lyophilized drug products and describes the issue with small datasets as follows:

*"ResNet-50 has about 23 million parameters, so training it on about 3000 images will just lead to the model memorizing all the images and their labels. By utilizing transfer learning, one seeks to train a large model with many parameters without having large amounts of training data."*

At the moment, rule-based algorithms are tailored to specific AVI equipment and even to specific products, showing limited generalization capability (18). Transfer learning of low-level features with good generalization across similar products may make it viable to use bracketing strategies during qualification for some models and to significantly reduce errors and model training time (18).

Veillon R et al also gave an outlook on the possibility of using automated machine learning "AutoML" to further automate the process of vision recipe development. The AutoML pipeline covers data preparation, feature engineering, model generation and model evaluation (71). This research field is still in its infancy and a clear understanding of AI/ML, trust and technological maturity needs to be established before reducing or removing the manual element in vision recipe development (18). However, fast technological advancements in chip manufacturing, available compute power and the field of AI/ML in general may accelerate the availability of such "hands-free" technologies. Large players in the AI-space already do have AutoML services for computer vision in their portfolio, for example *Google Cloud AutoML* (72), *Amazon SageMaker Autopilot* (73) and *Microsoft Azure AI Custom Vision* (74). Clear regulatory requirements and guidelines for AI/ML as well as GxP-compliant software services from vendors will be needed to find this automation technology in pharmaceutical manufacturing.

### 4.1.3 Use case 3: Gaining data-driven insights into the manufacturing processes

In 2006 the renowned data scientist Clive Humby coined the phrase: "Data is the new oil" (75). Just as crude oil requires refining to unlock its value, raw data must be analyzed, processed, and analyzed to generate meaningful information. This analogy highlights the value of data in gaining actionable insights. In AVI, a huge amount of data is generated during 100% inspection since different camera stations capture images from different angles and with different illumination setups or imaging techniques (76). Information extracted from retained images of 100% inspection may be an untapped data-source that can be used for root-cause-analysis for deviations, predictive maintenance or as source for process parameters when using digital twins in a holistic control strategy (77).

Yadav V and Kenney C describe the use of AI/ML on data from AVI as a business tool to learn about the rejected containers and to understand the root cause of false rejections to take corrective action (78). Rather than using AI/ML as part of the AVI system during the GMP inspection process, a ResNet50 CNN model was used as a diagnostic tool to extract information from the vast amount of images already available to them. Notably the authors write:

*"If a convolutional neural network could be developed to identify signals buried within images, it could reveal insights into the process that cannot be seen by humans. Given the breadth of analysis possible with AI/ML, existing vision recipes could potentially be drastically improved by incorporating models developed from large classified sets of images instead of relying on the traditional inspection window techniques currently in use."*

The goal for the AI/ML "hawk AVI" solution presented by the authors was used to enhance intra-batch troubleshooting and to gain a deeper understanding of rejects that can be used for AVI machine tuning. Data on component-specific performance associated with product rejection should also be generated to facilitate discussions with suppliers, refer to **FIGURE 5**. Furthermore, trending and insights into root-causes should be provided for upstream manufacturing operations for intra- and inter-batch troubleshooting. To achieve these goals, access via an user interface tool for the vision inspection team as well as for upstream process teams was proposed to facilitate root-cause-investigation and a rapid-response following the GMP process (79). Access to component specific information may also be used to gain insights into changes of feature design spaces that are caused by natural variability of materials, unintended changes by suppliers or differences when the same components are sourced from different material suppliers.

Data-driven approaches to root-cause-analysis were also described by Straub J in a 2019 presentation called "Collection, Archival, and Analysis of Images from Automated Vision Inspection: Present Value and Prospective Opportunities" at the PDA Visual Inspection Forum (80). The case studies presented by Straub showcased the opportunities of using existing data such as from image retention databases ("defect libraries") from AVI machines to gain insights for AVI process improvements. The case studies revealed the influence of material variations such as vial heel radius, bottom profile, shoulder profile, surface finish of crimping caps, as well as the root-causes for rejection due to camera failures, contamination in the background of images that can be seen through the glass vial or light reflections on surfaces. In his presentation, Straub concluded to apply DL to archived image libraries to drive AVI process improvements and to further reduce false rejects.

Unexpected or previously unseen defects may also be uncovered by DL anomaly detection using unsupervised learning (18). Rather than training a model to identify specific defects, the model is only trained on what good products look like which enables assigning an anomaly index to images or regions of regions. A high anomaly index would suggest that an image is unusual. A thorough understanding of

what an acceptable product looks like is essential when mapping the design space since the count of possible outcomes approaches infinity for processes that are not deterministic (81,82).

### 4.1.4 Summary (RQ1)

In summary, the literature indicates that applying AI/ML techniques to the visual inspection (VI) of parenteral drug products may reduce inspection errors, accelerate the development of vision recipes, and help to gain insights from upstream processes. Each use case demonstrates both operational benefits and technical challenges. For instance, studies by Tsay and Li (2019) and Zhao et al. (2018) report improvements in defect classification accuracy through DL, yet common limitations such as small datasets and issues due to variability in product appearance remain. These findings suggest that while AI/ML can significantly enhance inspection processes, further research is needed to optimize data augmentation strategies to ensure robust performance.

The promising capabilities observed in VI applications naturally lead to an examination of the broader quality management challenges that accompany AI/ML integration in GMP environments (RQ2).

## 4.2 AI/ML in GMP: Challenges and Opportunities (RQ2)

Depending on the use case, AI/ML may be deployed as a standalone software or as a sub-system (module) within a traditional software package. As AI/ML technology is not yet widely used in pharmaceutical manufacturing, this section explores aspects that should be considered in a GMP environment when viewed from the perspective of quality assurance. The challenges and opportunities examined in this section focus on the intended use as an AI/ML sub-system in automated visual inspection. However, many aspects do apply to the use of AI/ML software in general. The following quality management system (QMS) concepts will be used to structure this section:

- **4.2.1 Risk Management**
- **4.2.2 Data Governance**
- **4.2.3 Change Management**

### 4.2.1 Risk Management

Principles of a risk-based approach are crucial to assess, control and review risks from using AI/ML software to ensure patient safety and product quality. Robust frameworks such as the harmonized guideline ICH Q9(R1) Quality Risk Management are usually already in place in pharmaceutical QMS and its principles should be used to address specific AI specific risks as well. The importance of a risk-based approach for AI becomes evident when looking at the European Union harmonized rules on Artificial Intelligence, or “AI Act” (15), that regulate risk-classification of AI-systems depending on the intended purpose as unacceptable risk, high risk and low or minimal risk, see **Figure 13**.

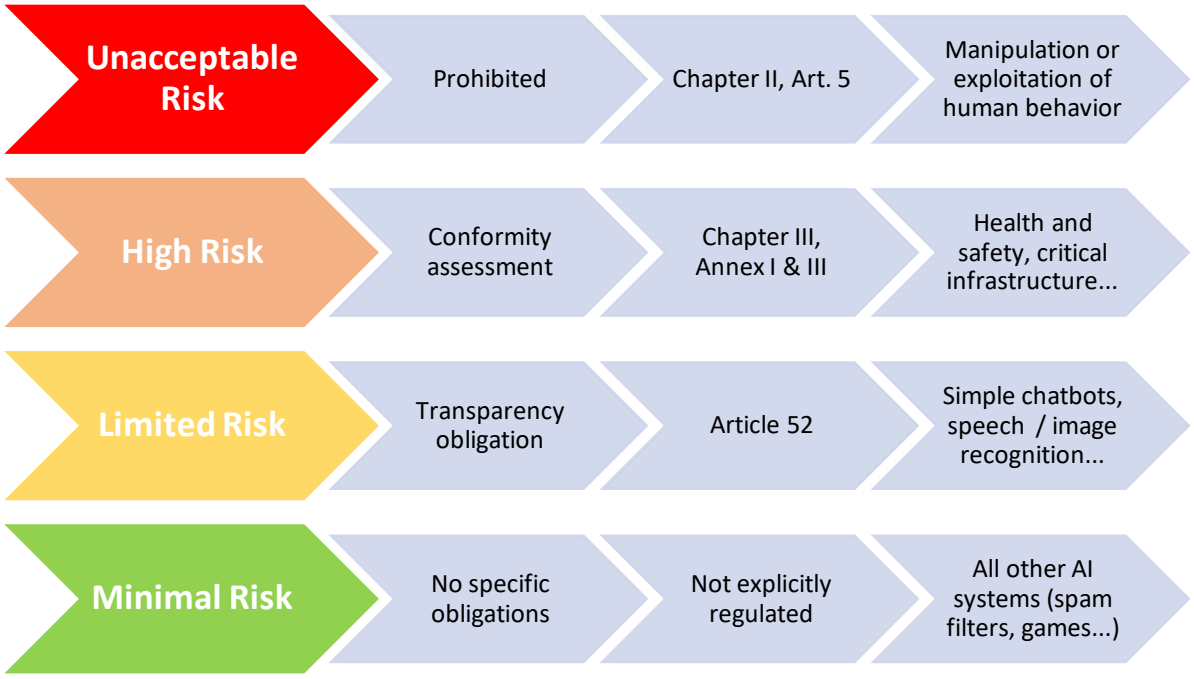


*Figure 13 EU AI Act risk types classification (15). From left to right: Risk classification of the AI-system, obligations, reference to the EU AI act chapters or articles, and examples of AI-systems within the risk category.*

Unlike the EU, the United States of America does not yet have a single, comprehensive federal law dedicated to regulate AI. The National Institute of Standards and Technology (NIST) has released the Artificial Intelligence Risk Management Framework "AI RMF" (83) and executive order 14110 "Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" was signed by US President Joe Biden October 2023 to mitigate risks associated with this new technology. The Executive Order was rescinded by President Donald Trump in January 2025, refer to section **4.3.2 REGULATORY INITIATIVES BY FDA**.

Since no sector-specific regulatory guidance for the use of AI/ML for medicinal products manufacturing is in place, guidance documents from pharmaceutical industry associations emerged. In October 2020, the International Society of Pharmaceutical Engineering (ISPE) published ISPE GAMP® RDI Good Practice Guide: Data Integrity by Design (84) including "Appendix S1 – Artificial Intelligence: Machine Learning" addressing the importance of data and data integrity. Around two years later in July 2022, the second edition of ISPE GAMP 5: A Risk-Based Approach to Compliant GxP Computerized Systems (85), adding "Appendix D11 – Artificial Intelligence and Machine Learning (AI/ML)", was published. Although not legally binding, this guideline is widely used in the pharmaceutical industry when it comes to the implementation of GxP-compliant computerized systems. A GAMP 5 lifecycle model for AI/ML sub-systems consisting of a concept-, project-, and operational phase is included in the guideline.

Blumenthal R et al (86) propose the "Machine Learning Risk and Control Framework" as a holistic approach to risk management to ensure quality and regulatory compliance within the life science industry. The approach is based on the ICH Q9(R1) process and combines previously published frameworks such as the AI maturity model (87), the GAMP 5 second edition AI/ML sub-system lifecycle approach (85,88), an AI Governance and QA Framework (89), and an adapted version of the Risk Analysis and Mitigation Matrix (RAMM) (90). The ML risk control framework has the following four major steps, that are embedded within the ICH Q9(R1) process, see **FIGURE 14**.

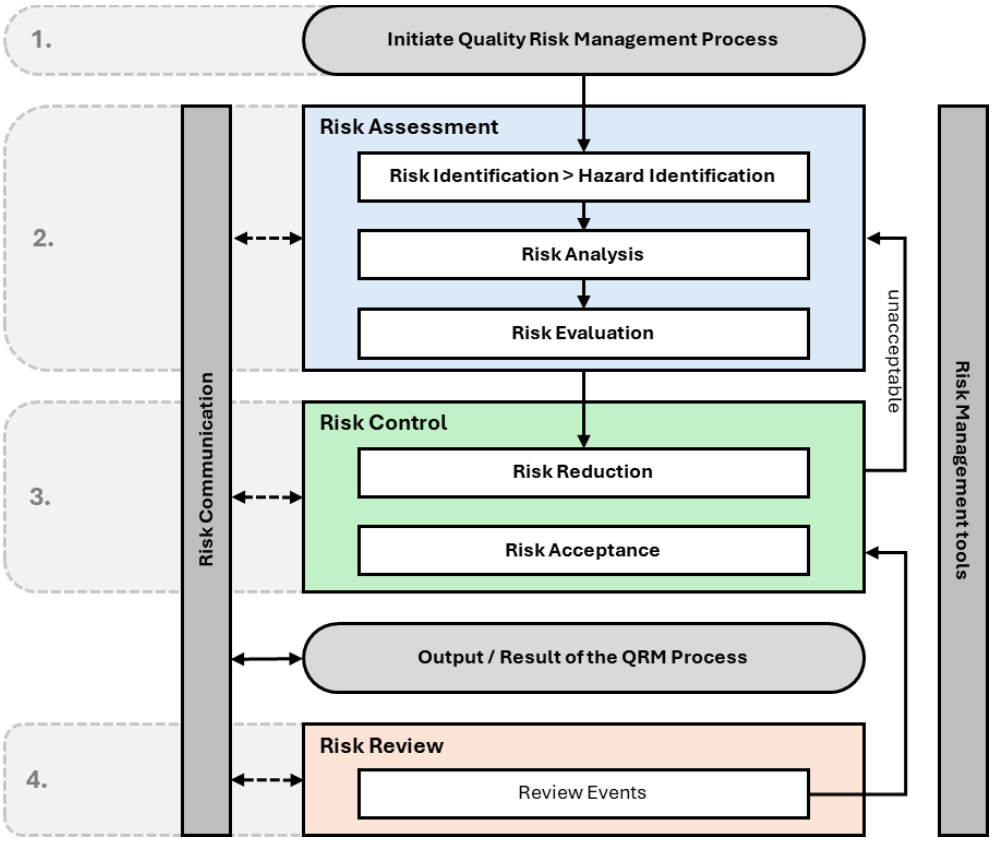


*Figure 14 Modified ICH Q9 Quality Risk Management (QRM) Process with references (1.-4.) to the ML Risk and Control Framework from Blumenthal et al (86,91).*

1. **Initiation of the QRM process**: The maturity level, see **TABLE 13**, of the AI-system is mapped against potential risk impact to patient safety. Through the intersection between AI-maturity level and the risk impact to patient safety, the overall *hazard (risk) impact factor* is determined, which will later on be used as an overall multiplicator in risk assessment.
2. **Risk assessment**: The GAMP 5 lifecycle model is used to identify hazard clusters and analyze the individual risks within these clusters along the major stages of concept-, project-, and operation phase. Examples for hazard clusters: Initial data set quality, data split, model design, model training, model evaluation, deployment and release, data quality in operation and human interaction and monitoring. As a risk management tool, "AI RAMM" is used instead of FMEA to combine the risk rating of the whole hazard clusters with the individual risks including color coding against critical quality attributes (CQAs), see **FIGURE 15**.
3. **Risk control**: Measures to reduce risk include additional testing, extra procedural or technical controls, and collecting more training data. Another options to reduce the risk of an AI-system may be achieved to adjust the maturity level of the AI-system via its autonomy or control design, see **TABLE 13**.
4. **Risk review**: Periodic review of risks documented in RAMM.

Next to identifying the risk impact factor, the AI-system maturity level can be also used to identify the necessary depth of CS validation. Erdmann N et al (87) as part of the ISPE D/A/CH Working Group on AI Validation proposed the AI Maturity Model for GxP Applications as the foundation for AI Validation, addressing the lack of missing AI validation concepts. The concept determines the "AI system maturity" and the corresponding depth of validation needed based on the combination of control design and its autonomy.

The control design defines different stages by which the AI-system can *take over controls that influence product quality and safety*. The spectrum of the control design ranges from AI-systems being used in parallel to GxP process and have no direct influence on decisions (stage 1) to AI-systems running automatically with the ability to independently initiate changes to correct itself (stage 5).

The autonomy of an AI-system refers to the capability of automatically updating itself by retraining. Fixed algorithms without ML (stage 0) as well as AI-systems in a locked state with manual retraining (stage 1) all the way to fully automated AI-systems that self-determine its task competency and strategy (stage 5).

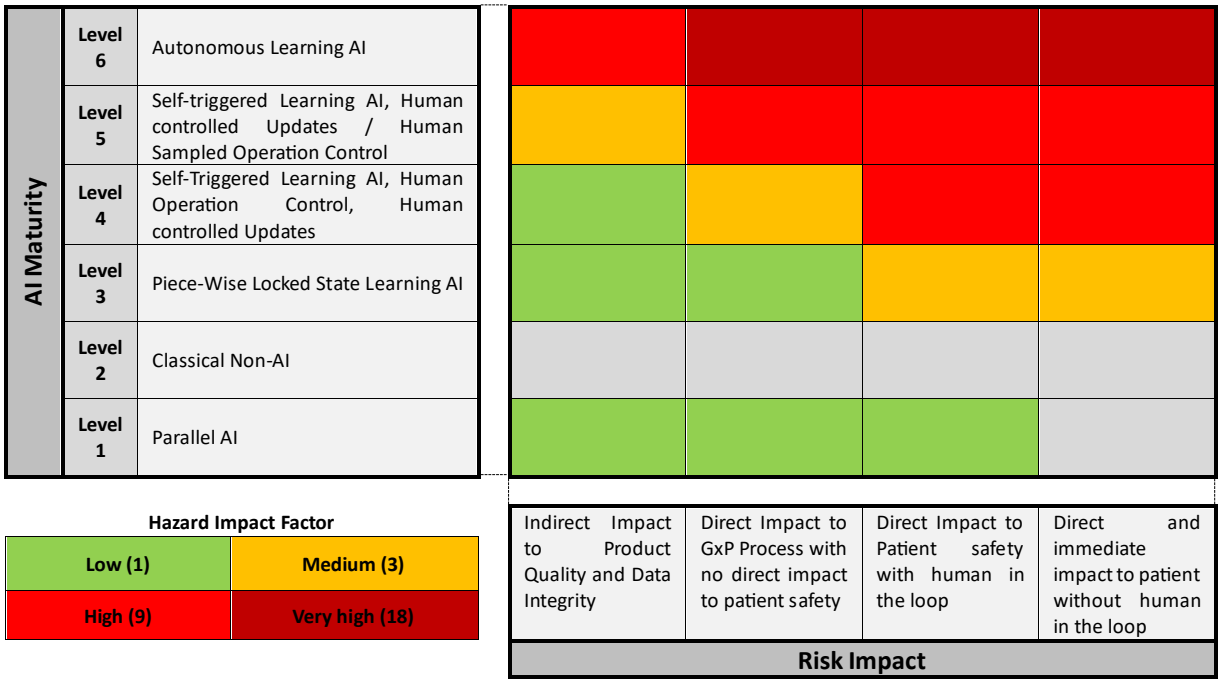


*Table 13 Adapted from Source: ML Risk and Control Framework – "Figure 4: Risk severity matrix defining the hazard impact based on the AI maturity level and the risk impact", © PE magazine Jan/Feb 2024 (86).*

**Hazard Impact Factor (HIF)** 9

| Quality Dimension | | Data Quality Mgmt | | Human-AI Interaction | | Predictive Power | | Stability | | Calibration | | Total Score | Hazard Cluster Total Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hazard Cluster | Risk | RA | RA*HIF | RA | RA*HIF | RA | RA*HIF | RA | RA*HIF | RA | RA*HIF | | |
| Initial Dataset Quality | Risk 1 | Low | 9 | Low | 9 | Low | 9 | Low | 9 | High | 81 | 117 | 198 |
| | Risk 2 | Low | 9 | Low | 9 | Low | 9 | Medium | 27 | Medium | 27 | 81 | |
| Data Split | Risk 1 | High | 81 | Low | 9 | Low | 9 | Low | 9 | Low | 9 | 117 | 261 |
| | Risk 2 | Medium | 27 | Low | 9 | Low | 9 | Low | 9 | Low | 9 | 63 | |
| | Risk 3 | Low | 9 | Low | 9 | Medium | 27 | Low | 9 | Medium | 27 | 81 | |
| Model Design | Risk 1 | Medium | 27 | Medium | 27 | Medium | 27 | Medium | 27 | Medium | 27 | 135 | 135 |
| Model Training | Risk 1 | High | 81 | Medium | 27 | Medium | 27 | Low | 9 | Low | 9 | 153 | 153 |
| Model Evaluation | Risk 1 | Low | 9 | Low | 9 | High | 81 | Low | 9 | Low | 9 | 117 | 117 |
| Deployment and Release | Risk 1 | Medium | 27 | Low | 9 | Low | 9 | Low | 9 | Medium | 27 | 81 | 81 |
| Data Quality in Operation | Risk 1 | Low | 9 | Medium | 27 | Low | 9 | Low | 9 | Low | 9 | 63 | 63 |
| Human Interaction and Monitoring | Risk 1 | Medium | 27 | High | 81 | Low | 9 | Low | 9 | Low | 9 | 135 | 135 |

*Figure 15 AI RAMM risk management tool from Blumenthal et al* (86). *Individual risks are assessed through different quality dimensions and multiplied with the overall Hazard Impact Factor (HIF) to determine the overall risk score of the hazard cluster.*

Depending on the categorization of the "validation level", additional quality assurance activities and requirements compared to traditional computerized system validation (CSV) are proposed. More validation activities and requirements for the computerized system (CS) are proposed by the authors with each level adding further activities. A modified table from the article by Erdman N et al (87), that summarizes the validation levels and the associated activities derived from the control design and the autonomy of the AI-system, can be seen in **TABLE 14**. The examples of AI-specific activities and

requirements listed by the authors serve as a good starting point for risk-based approaches for validation, dependent on the intended use of the AI-enabled computerized system.

| CS validation level description | | Minimum Validation Activities and Requirements |
|---|---|---|
| I | Parallel (AI) CS | No validation required |
| II | Classical non-AI CS | Validation of CS, but no dedicated focus on AI |
| III | Piecewise locked state AI CS | In addition to the above requirements:<br>- Documented justification on why a model was selected<br>- Training data verification<br>- Model quality assurance after training<br>- Input data monitoring in operation<br>- Retraining procedures defined |
| IV | Self-triggered learning AI CS with human operation and update control at all times | In addition to the above requirements:<br>- Monitoring of model quality in operation<br>- Controlling quality key performance indicators (KPIs) and notification process<br>- Validation of the human factors (depending on control design) with regards to overrides, qualifications, and AI system acceptance |
| V | Self-triggered learning AI CS with update control, but overall or sampled operation control only | In addition to the above requirements:<br>- Periodic re-test with defined test data set<br>- Assurance of self-control<br>- Control of AI system outcomes by samples for a defined, risk-oriented fraction, and adequate stratification of input/output instances |
| VI | AI CS with autonomous learning | *Validation concept currently under development* |

*Table 14 Adapted from Source: AI Maturity Model for GxP Application: A Foundation for AI Validation – Table 3: Validation levels, © PE magazine Mar/Apr 2022 (87).*

Veillon R (92) presented a validation approach in a case study using supervised deep learning in AVI during the FDA/PQRI workshop on "Regulatory Framework for the Utilization of Artificial Intelligence in Pharmaceutical Manufacturing". The specific use case for an edge infrastructure included an AVI-system, an image server database, a webserver, a labelling tool, training software, and a digital twin. The User Requirements Specification (URS) in the case study covered the image database structure including metadata, the image storage and retention application, image labelling and training, as well as a digital twin app. The URS was accompanied by the network landscape definition, a system risk assessment, servers (test, validation, production) and concluded by operational and performance qualification (OQ-PQ). Risks related to image processing and the AI-model were included in the system risk assessment with the aim to ensure data integrity, product quality and patient safety. A desktop OQ on an edge digital twin with a focus on metrics and its documentation was described by Veillon, followed by AVI hardware testing to test digital images. After OQ, a real defect kit was tested, and the results were compared to the baseline of human MVI ("gold standard") and thereafter production lots. The validation approach also included a continuous life-cycle improvement approach embedded in a change control process, where gathered data is used for retraining the model followed by re-validation. On a more specific note, Veillon R proposed information to be included in OQ documentation, see **TABLE 15**, and concludes that validation should rely on a strong URS, QRM, extensive process understanding, and transparency to enhance the explainability of AI.

| Model training | OQ documentation |
|---|---|
| **Input** | Description of the task of the model – for example: classification, localization or segmentation |
| | Description of the AI framework used (TensorFlow, Torch, Keras, Caffe, Theano, MXNet…) |
| | If the model has been trained from scratch or if transfer learning has been used |
| | List of images used in the model training data sets as well as the composition and balance of the training / validation / test set. |
| | Description of the hyperparameters used. For example:<br>- Network architecture (number, type and size of layers)<br>- Learning rate (step size during weight updates)<br>- Batch size (number of training samples per iteration)<br>- Number of epochs (number of passes through the dataset)<br>- Optimizer (algorithm to update the model weights)<br>- Loss function (measures error between predictions and actual values)<br>- Regularization (techniques to prevent overfitting) |
| | Training data: Description of the labelling software used and the personnel involved in labelling. |
| | Description of data augmentation and, if used, use of Generative adversarial networks (GAN) to generate data. |

| Model training | OQ documentation |
|---|---|
| | Input image pre-processing |
| **Output** | Name and version of the trained network |
| | Documentation of learning curve with loss |
| | Performance metrics related to confusion matrix. For example: accuracy, precision, recall, F1-score… |
| | Documentation of the heatmap visualization for test set images |
| | Performance evaluation with images from a test kit never seen before |
| | Evaluation of false rejects on larger sample sets. |

*Table 15 Documentation of model desktop OQ. Appended information based on the presentation from Veillon R (92)*

#### 4.2.1.1 Risks related to the AI/ML-model and data

AI-specific risks related to the model and the data used for training as well as the suitability of the model itself for its intended use should be covered by the validation documents. Also risks after the development phase is completed need to be considered. For example, when using an AI-system that was developed using supervised training and put into a locked state, the input data needs to be monitored in order to detect if the statistical distribution of the input data has changed in a way that negatively impacts model performance for its intended use and to understand when retraining under change control procedure needs to take place. Operational controls through monitoring of performance parameters and metrics should ensure, that the system is operating within the acceptable limits and must include a mechanism to indicate when retraining is needed (93). In the case of Visual Inspection, verification of inspection results in the form of a confusion matrix may be an option, refer to **TABLE 8** and **TABLE 9** . Performance metrics are important to evaluate if the systems works as intended and therefor drive the iterative improvement process inherent to AI/ML (88).

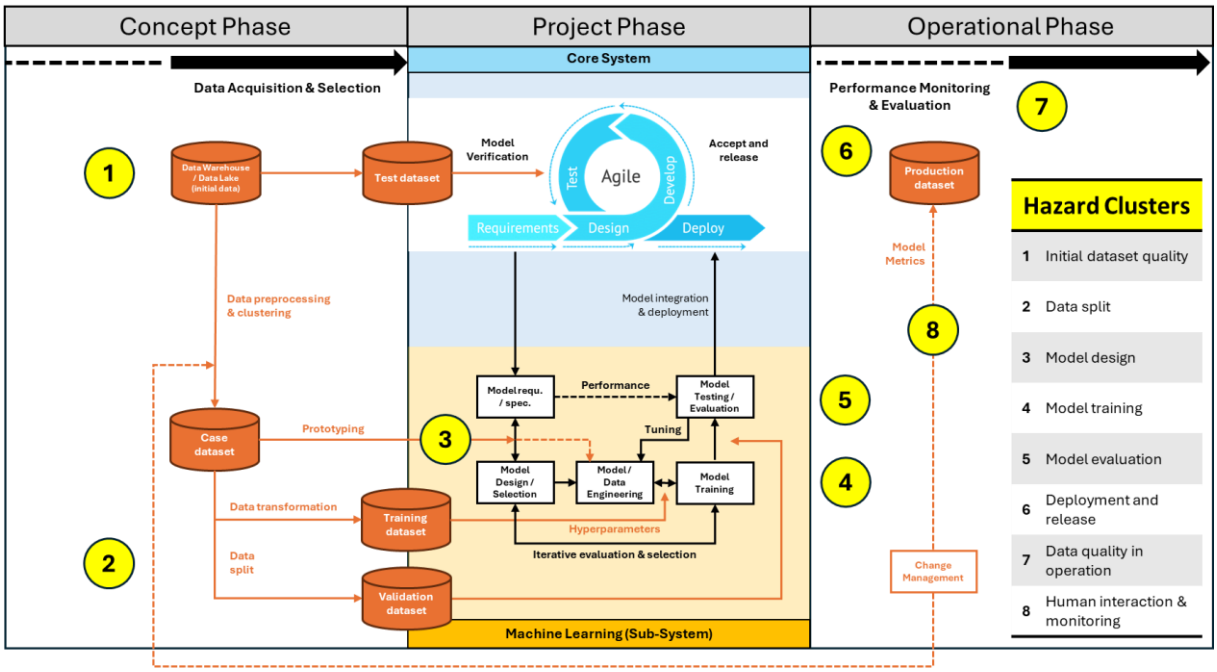


*Figure 16 Adapted overview of the GAMP 5 ML lifecycle with associated hazard clusters based on Blumenthal et al (86).*

When compared to "traditional" rule-based software, new risks related to the software design principle need to be assessed and controlled, see **FIGURE 16**. The ISPE GAMP5 guideline underscores several key aspects for AI/ML software development. It emphasizes that ensuring data integrity from the concept phase is critical, thus needing robust data management practices to safeguard overall quality. Additionally, the guideline recognizes the iterative nature of developing AI/ML sub-systems within IT infrastructures, advocating for agile methodologies as a more suitable alternative to the traditional V-model approach. It further highlights the importance of incorporating performance metrics into

technical specifications, encouraging a data-centric development paradigm, and instituting proactive risk management.

#### 4.2.1.2 Risks related to External Software Suppliers, IT Infrastructure, and Cybersecurity

Implementing AVI with AI/ML a sub-system requires robust software solutions and reliable IT infrastructure. These systems often rely on commercial off-the-shelf (COTS) software or "X-as-a-service" models, which introduce additional risks concerning supplier reliability, data handling, system updates, and data oversight. Moreover, cybersecurity vulnerabilities can arise from both external service providers and internal IT infrastructure, highlighting the need for a comprehensive risk management strategy.

When manufacturers opt to use COTS software or SaaS, they become dependent on external vendors for critical aspects of the AI/ML system, including model updates, hyperparameter settings, and data handling protocols. According to FDA guidance (94), appropriate oversight of vendors and data storage solutions is essential to ensure the validated state of GxP systems is maintained. In practice, quality and confidentiality agreements may be necessary to guarantee transparency about the vendor's AI-model architecture, version control, and system updates.

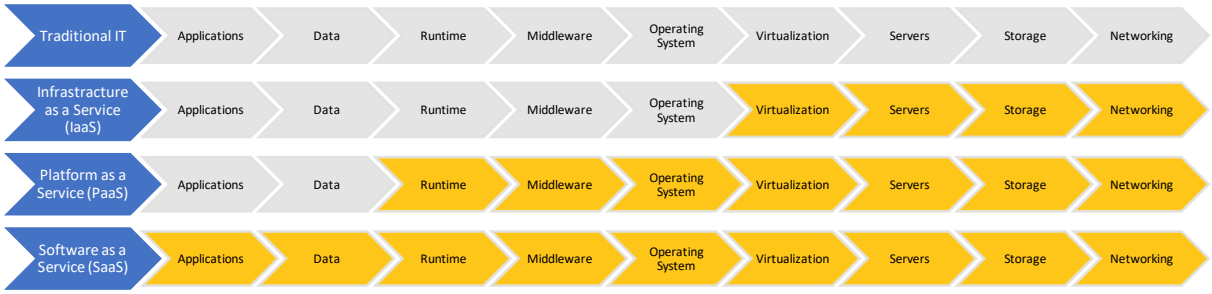


*Figure 17 Different XaaS service models compared to traditional IT-infrastructure. Grey color = self-managed, orange color = supplier managed (95).*

**FIGURE 17** illustrates how responsibilities for applications, data, runtime, middleware, and infrastructure differ among various "XaaS" models compared to traditional on-premise IT setups (95). The maturity of the supplier's IT infrastructure and security policies directly affects the integrity and reliability of the AI/ML system. In addition, many software vendors may not be fully familiar with GMP requirements. Consequently, formal training, clear documentation of activities, and thorough supplier management becomes essential.

Usage of open-source AI solutions is also widespread in AI/ML (Meta Llama, Deepseek base models,…), necessitating discussions on licensing, data privacy (GDPR compliance), and security. A multi-disciplinary team of subject matter experts (including data scientists, software engineers, legal - and compliance officers, as well as quality assurance personnel) should therefore be involved in framing requirements, deliverables, and ongoing vendor oversight.

Technical limitations of the IT infrastructure can undermine the performance and reliability of AI/ML subsystems for AVI, especially for high-throughput environments where fast accept/reject decisions are crucial. For instance, edge computing can minimize latency by processing data directly on local devices; however, it also introduces additional risks and complexity in data synchronization. Yadav et al. (78,79) highlighted common pitfalls such as:

- Some AVI systems do not retain images on the local edge computer.
- Large image sets require scalable infrastructure to handle near real-time inference, storage, and query functionalities.

- Agile software development practices are vital for iterative improvements of AI/ML models, which deviates from the V-shaped model traditionally used in CSV.

Cloud computing provides vast storage and computational resources, but intermittent availability and reliance on third-party data centers (96) can disrupt model training, corrupt datasets, or omit critical information. For real-time operations, balancing cloud-based and edge-based processing is key to maintaining both performance and data security. As AI systems often involve local servers, virtual machines, and cloud services simultaneously, the attack surface for cyber threats expands significantly.

The ISPE GAMP Good Practice Guide on IT Infrastructure Control and Compliance (95) offers guidance on meeting regulatory expectations in GxP-regulated environments. While not explicitly tailored to AI/ML, it addresses critical topics - such as virtualization, cloud computing, third-party data centers, and outsourcing - that are directly relevant to implementing AI/ML systems. Manufacturers must maintain a "state of control" over their IT infrastructure to protect the validated status of GxP applications. Key aspects include:

- Supplier assessment and management
- IOQ (Installation and Operational Qualification) of infrastructure components
- Configuration management and change control
- Risk management for IT infrastructure
- Involvement of service providers in critical infrastructure processes
- Service-level agreements (SLAs) with XaaS providers
- Security management (access controls, availability, data integrity)
- Data storage (security, confidentiality, privacy)
- Backup, restore, and disaster recovery
- Archiving and retrieval processes

In the discussion paper on "Artificial Intelligence in Drug Manufacturing," the FDA (94) emphasizes that cloud and edge computing can affect oversight of pharmaceutical manufacturing data and records. Third-party data management may extend beyond storage to AI-based analytics for process monitoring and advanced process control (APC). During inspections, the complexity of third-party services can pose challenges for ensuring proper data traceability, cybersecurity, and integrity. Therefore, quality agreements with third-party providers must explicitly address AI-related risks and responsibilities.

Major providers like Amazon Web Services (AWS) have been expanding their GxP service portfolios, issuing whitepapers and frameworks for ML best practices in healthcare and life sciences (97,98). Nonetheless, adopting cloud-based AI solutions demands rigorous controls to protect against unauthorized access, data corruption, or non-compliance with regulatory standards.

Cybersecurity concerns in AI extend beyond conventional software "bugs" or human coding errors. As the German Notified Bodies Alliance (IG-NB) (99) notes, AI-specific attacks exploit inherent algorithmic vulnerabilities, such as adversarial attacks on convolutional neural networks (CNNs). Minor changes or noise introduced into an image can alter an AI model's classification output (**FIGURE 18**), raising questions about the reliability of AI-driven inspection systems. Research from Su et al. (100) demonstrates that one-pixel attacks can mislead deep neural networks (DNNs), although the impact on pharmaceutical AVI may be limited in highly controlled environments. Nevertheless, maintaining awareness of such threats is crucial to ensure data integrity and protect against potential manipulations of input images.

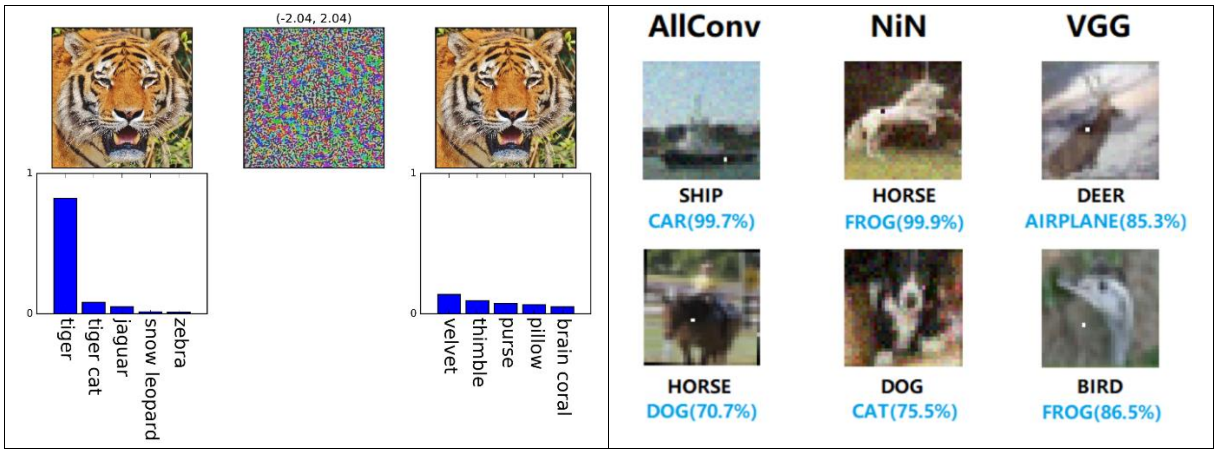


*Figure 18 Left side: Adversarial attack on VGG16 CNN model by adding noise on top of the original image (source: (44)). Right side: One-pixel attack on different images classified by different DNNs. Source: (100)*

Another significant example is the 2013 "Xerox-scandal" involving a bug in the JBIG2 lossy compression algorithm (101), where scanned documents underwent unintended character substitutions (6 < > 8 or l < > i) due to symbol coding and in essence created tampered data. While this incident did not involve AI/ML, it underscores the importance of image-processing vulnerabilities in regulated contexts.

#### 4.2.1.3 Risks related to Explainability and Interpretability

Explainable AI or "xAI" describes processes and methods that enable users to understand why the model concludes with a certain prediction based on the input data. The "black box" aspect of AI-models needs to be already considered during model-selection and is usually a trade-off between performance and explainability. Different explanation types and how they are perceived by the end-user, as investigated by Herm L et al (102), should be considered, see **TABLE 16** and **FIGURE 19**.

| Explanation type | Description of the explanation type |
|---|---|
| How | Which input areas are relevant to the trained model. |
| Why | Which areas of the given input are relevant to the outcome of the prediction. |
| Why-Not | Which areas of the given input are not relevant to the outcome of the prediction. |
| How-To | How hypothetical adjustments of the given input would result in a (different) model prediction. |
| What-Else | Explanation by example; showing similar inputs and their respective predictions. |

*Table 16 Explanation types and corresponding descriptions based on (102)*

The study by Herm L et al looked at images from MRI-scans and rated the explanation types by the end-user and found that explanations on "why" the outcome was predicted and examples of "what-else" are important to understand AI/ML output.

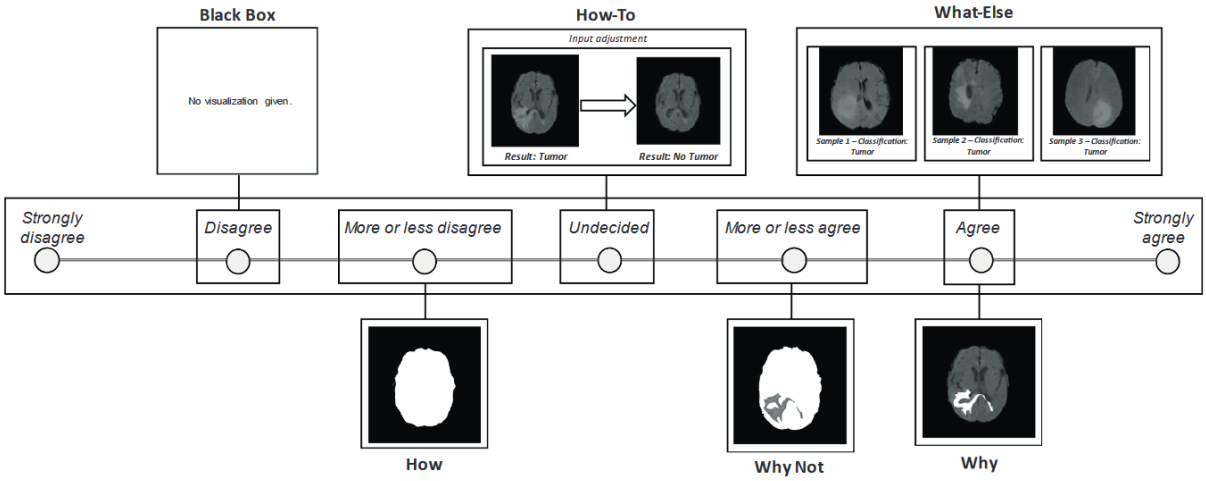


*Figure 19 Example of xAI on brain MRI-images with Illustration of different explanation types in MRI, along with end-users' ratings of each explanation type. By highlighting regions of the MRI or by providing examples for changes, the final output of AI becomes interpretable to the end-user. Image source (102)*

XAI features are necessary to improve the trustworthiness, reliability, and operational transparency of AI-driven visual inspection systems. The selection of explanation features should be dependent on the intended use of the AI/ML model. Tran P (68) used gradient weighted class activation mapping ("Grad-CAM") to produce visual explanations in the form of a heatmap of what regions of the input image were important for the models output. An example for Grad-CAM visualization is shown in **FIGURE 20**.

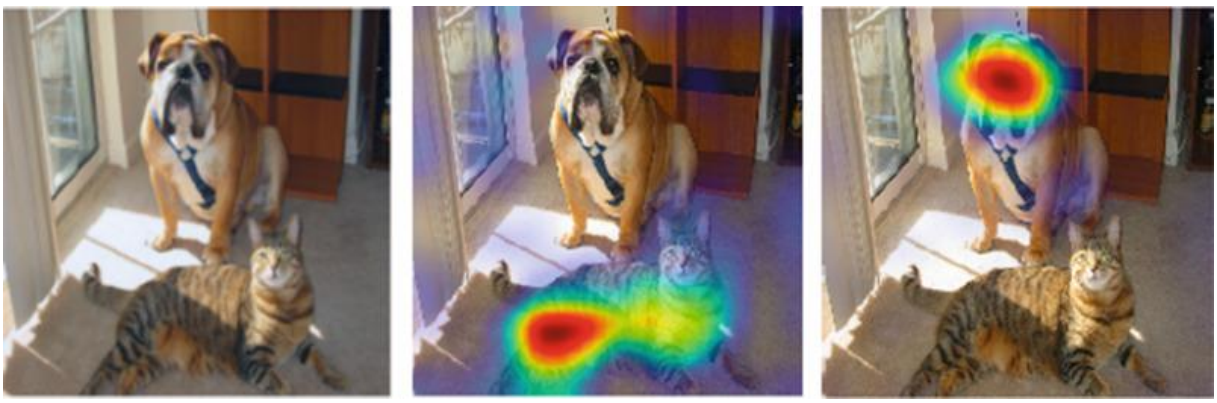

*Figure 20 Grad-CAM example from left to right: Original image; cat -, and dog saliency map overlay. Image source (103)*

For binary and multi-class classification models in VI, good visual explanations should be class discriminative and therefor be able to localize the output category within the image (103). Although not comprehensive, **TABLE 17** lists different xAI features that may be used in VI to shed light into the black-box aspect of otherwise opaque DL models.

| XAI feature | Description | Use case example for VI of parenterals |
|---|---|---|
| Feature attribution (saliency maps) | Highlights image areas that contributed most to the decision of the model. | Saliency map showing a crack or particulate contamination on the image of the container. |
| Class activation mapping (CAM) such as Grad-CAM. | A common type of saliency map used in computer vision. It finds regions that influence specific class predictions. | Helps identify which features differentiate between various defect types (bubble VS particulate matter). |
| Localization features such as Bounding Boxes (BBox) or segmentation | Precise localization of the defect in the image. | BBox to highlight all detected particulate matter. |
| Decision Confidence Scores | Provides a confidence score for each decision or classification. | Confidence scores to evaluate the certainty of the accept/reject decisions by the model. |
| Visual Counterfactual Explanations | Shows how changes in the image would change the model's prediction. | Visualizes how removing a defect, like cosmetic defects, would change the classification outcome. |

| XAI feature | Description | Use case example for VI of parenterals |
|---|---|---|
| Benchmark comparison | Compares current results with historical data or standards. | Shows how a detected defect compares with known issues in historical production data. |

*Table 17 Overview of key Explainable AI (XAI) features, their descriptions, and corresponding use-case examples in pharmaceutical visual inspection.*

**FIGURE 21** shows three examples of how xAI features for object detection problems may be used – images of arthropods and a cityscape is used for illustration purpose but the principle can be used on problems in visual inspection of parenteral drug products.

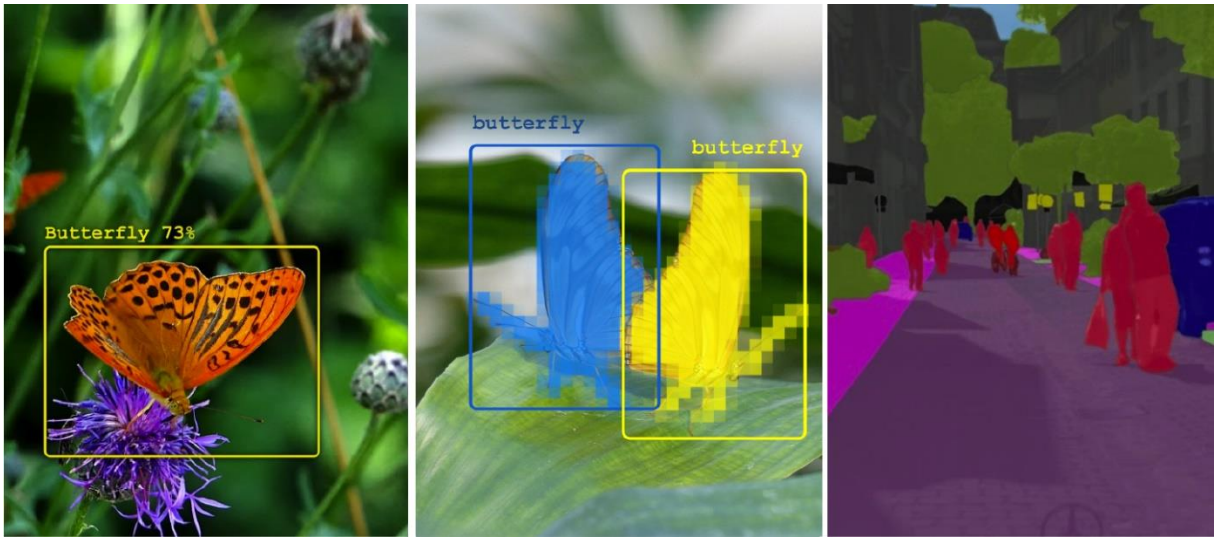


*Figure 21 Examples of xAI features for object detection problems in machine vision on images from Arthropods and Cityscapes dataset. From left to right: Object detection with BBox and a confidence score, instance segmentation, and semantic segmentation of the whole image. Image from Github repository of (104).*

The article "The Road to Explainable AI in GxP-Regulated Areas" from Altrabsheh E et al (105) discusses the increasing adoption of AI in regulated industries and the need for xAI to address the complexity of AI algorithms. Implementation of xAI as an additional module to provide human-readable explanations of AI decisions is essential for gaining trust and compliance with regulatory requirements. The black-box nature of AI presents significant challenges for stakeholders - such as personnel, quality assurance teams, and inspectors - who must verify that the system operates as intended to maintain both quality and safety.

The importance of explainability and interpretability is also expressed by EMA in section 2.4.5. of the draft “Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle”:

*“To allow review and monitoring, methods within the field of explainable AI should be used whenever possible. This includes providing feature importance lists, SHAP and/or LIME analyses or similar explainability metrics, both for the model and for individual inferences during deployment. Computer vision models, and extensions into other modalities where attention mechanisms are used, should whenever possible be supported by attention plots to verify that features are extracted from relevant positions in the image or sequence.”*

Furthermore, easy access to xAI features, for example through integration into the graphical user interface (GUI) of the software front-end, is essential for building trust and efficient human oversight. By offering transparent, understandable insights into AI-driven decisions, the GUI allows SMEs to check the decisions of the AI-system and verify the results with their expertise.

#### 4.2.1.4 Risks related to Regulatory Compliance

Regulatory risks may arise from several factors, including ambiguous expectations from national competent authorities during GMP inspections for validating computerized systems (CS) with integrated AI/ML sub-systems, uncertainties regarding data integrity, and challenges related to explainability, interpretability, and overall acceptance of this emerging technology. As many pharmaceutical companies produce for the global drug market, compliance to the AI-regulations of different markets is another risk that may need to be taken into consideration (106). A mitigation strategy could be a compliance monitoring system led by a regulatory intelligence team that maintains processes for staying updated on regulatory changes (107).

With increasing automation and complexity of computerized systems, the impact of AI-systems on the roles and responsibilities on individuals in key roles, such as the Qualified Person (QP) in the EU, within the manufacturing process is another risk to organizations (108).

Mintanciyan A et al (109) described a framework in their article "Artificial Intelligence Governance in GxP Environments" detailing AI-specific risks that should be covered by overarching AI governance policies that are operationalized by procedures. The authors conclusion emphasizes that effective AI governance requires collaboration across different organizational levels to establish an AI governance framework to ensure GxP-compliance, thus preventing and resolving potential quality and data integrity issues, see **TABLE 18** for an example of policies and procedures in an AI framework.

| | Policies | Procedures |
|---|---|---|
| **AI Governance** | Data Privacy and Protection Policy | Data management procedure |
| | Fairness and Non-Discrimination Policy | Bias detection and mitigation procedure |
| | Transparency and Explainability Policy | Model development and validation procedure, Transparency and explainability procedure |
| | Accountability and Responsibility Policy | Roles and responsibilities for developing, deploying and managing of AI-systems. Procedures for AI-related incidents |
| | Security and Safety Policy | Procedures for vulnerability assessment, secure coding practices, and response to security incidents |
| | Human-AI Interaction Policy | Procedures for human-oversight and intervention, handling of errors |
| | Continuous Monitoring and Auditing Policy | Procedures for tracking the performance of the AI system via metrics/KPIs and procedures to initiate model re-training |
| | Stakeholder Engagement Policy | Procedures for engagement of internal (employees) and external (regulators, vendors) stakeholders |
| | Training and Education Policy | Procedures for training on AI governance, ethics, and best practices. |
| | Legal and Regulatory Compliance Policy | Procedures for ensuring regulatory compliance. |

*Table 18 Examples for AI Governance policies and procedures based on Mintanciyan A et al (109)*

In February 2021, the medicines inspector Ib Alstrup of the Danish Medicines Agency (DKMA) was among the first to publish questions related to the use of AI/ML in critical GxP applications in Europe (58), refer to section **4.3.4.1** for details. In August 2022, the German health inspectorate ZLG expert group for computer-aided systems (EFG 11) published their "Aide-Mémoire 07121203" on inspection of computerized systems (110), refer to section **4.3.4.2**.

Good AI/ML documentation practices are necessary for regulatory inspections, providing inspectors with the necessary information to evaluate compliance with regulatory guidelines. Such documentation can enhance the transparency of AI/ML processes, ensuring that all data used is traceable and meets regulatory expectations. However, potential knowledge gaps may arise in understanding complex data science issues, statistics, and AI/ML topics which can hinder evaluation and oversight of AI/ML systems. Bridging these gaps through training and upskilling initiatives is crucial to ensure manufacturing personnel have the capability to assess AI/ML applications adequately.

Additionally, multidisciplinary teams, including data scientists, statisticians, quality assurance experts, and IT personnel, may be necessary during inspections to reconstruct and interpret inputs and outputs of AI/ML models. Such teams can help ensure that all aspects of the AI/ML models, from data integrity to model outcomes, are thoroughly evaluated and understood.

### 4.2.2 Data Governance

Data governance is a critical foundation for AI/ML model training, particularly in the pharmaceutical industry, where data quality can directly impact the safety of drug products. AI/ML models are intrinsically data-driven, as they extract or adapt their parameters from training data. In this context, the training data serve as the “raw material”, the algorithm represents the “manufacturing process”, and the trained AI/ML model is the “finished product”. Hence, the quality of data used for training affects the predictions of the software –"garbage in, garbage out" or "bias in, bias out". Governance and a clear data management strategy are necessary as a foundation for AI/ML model training and need to be extended to the concept phase of AI/ML projects to avoid human bias and to ensure data integrity and traceability.

#### 4.2.2.1 Data Integrity Guidelines

The 2024 final EMA CHMP/CVMP joint scientific guideline on "The use of Artificial Intelligence (AI) in the medicinal product lifecycle" (52) emphasizes that the inherently data-driven approach necessitates active measures to minimize the integration of bias into AI/ML applications. Section "2.6. Governance" of the scientific guideline suggests the implementation of SOPs to cover the following high-level aspects:

*“SOPs implementing GxP principles on data and algorithm governance should be extended to include all data, models and algorithms used for AI/ML in cases of high regulatory impact or high patient risk. Aspects related to the governance of all components used, the application of data protection and compliance with applicable data protection laws and ethical standards should be documented and regularly reviewed.”* (52)

In the case of pharmaceutical manufacturing, GxP principles on data integrity can be found in several different guidance documents by regulatory bodies and industry organizations, see **TABLE 19**.

| Issuing Body / Org. | Document title | Scope | Reference / Date |
|---|---|---|---|
| PIC/S | Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments | Global standard on data integrity in GMP/GDP environments | (111) / 01.07.2021 |
| US FDA | Data Integrity and Compliance With Drug CGMP: Questions and Answers | Data integrity and compliance in US GMP environments | (112) / 13.12.2018 |
| US FDA | Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations | US clinical investigations Q&A. Guidance for industry | (113) / 02.10.2024 |
| EMA | EU Good Manufacturing Practice (GMP) Volume 4, Part I - Basic Requirements for Medicinal Products: Chapter 4 - Documentation | EU GMP Guideline for medicinal products. | (114) / 30.06.2011 |
| EMA | Guideline on computerised systems and electronic data in clinical trials | EU clinical trial guideline on computerized systems and electronic data | (115) / 09.03.2023 |
| EMA | Guidance on good manufacturing practice and good distribution practice: Questions and answers | EU GMP FAQ on data integrity by GMP/GDP Inspectors Working Group | (116) / online |
| UK MHRA | GXP Data Integrity Guidance and Definitions | UK-specific guidance | (117) / 09.03.2018 |
| WHO | Annex 4 Guideline on data integrity | Global standard | (118) / 10.10.2021 |
| ISPE | Good Practice Guide on “Data Integrity by Design” | Industry best practice guide | (84) / 10.2020 |

*Table 19 Overview of key data integrity guidance documents from various regulatory agencies and industry bodies*

These documents provide the necessary principles to maintain data integrity throughout the lifecycle of AI/ML models. Guiding principles such as "ALCOA," or its extended version "ALCOA++," remain

relevant for AI/ML to ensure that data is attributable, legible, contemporaneous, original, accurate, as well as consistent, enduring, available, and traceable (119). For example, implementing ALCOA principles can ensure that each piece of data used in AI/ML is documented in a way that makes it traceable back to its origin, helping to prevent issues related to data integrity.

#### 4.2.2.2 Datapoints and Datasets

AI/ML models may require vast amounts of data to develop an algorithm with satisfactory generalization capabilities, and this raises the issue of inadequate dataset quality. In an effort to improve transparency, accountability, and outcomes of AI systems, recent research papers have proposed new concepts such as "The Dataset Nutrition Label" (120), "Datasheets for Datasets" (121), and "Data Cards" (122) to systematically describe, summarize, and document the data used. These tools serve as an innovative way to provide key information about datasets, including their composition, potential biases, and intended use. The term “Data Card” can also be found in the “FDA Digital Health and Artificial Intelligence Glossary – Educational Resource” (51).

The critical role of data and its technical aspects such as data acquisition and augmentation are described in section "2.5. Technical aspects" of the EMA reflection paper (52). In the case of a dataset for training of a supervised CNN multi-class model to categorize vials of a drug product (accept, particle, fibers, etc.), the following points may be considered for the characterization of the dataset:

- **Dataset snapshot**: Version and versioning approach of the dataset, audit trail for datasets and datapoints, number of unique images, inclusion and exclusion criteria for data points (Probability of Detection “PoD” and ground truth), number or percentage of synthetic data in the dataset, storage location of the dataset (repository, local folder, cloud, etc.).
- **Metadata about image acquisition**: Who, when, where and how?
    - Image properties in the dataset: Resolution, color space (RGB), filesize, image format, compression algorithm (lossy, lossless),
    - Traceability to a physical object: specific batch of a product or to a specific real-world test-kit.
- **Data augmentation**:
    - Techniques used: Geometric transformation such as flipping or sheering, addition of noise, change of contrast, brightness, color, resolution…
    - Rationale, software and workflow used for augmentation…
- **Feature design space**: description of the different classes, handling of ambiguous edge cases, handling of unknown/new defects…
- **Data cleansing strategy**: Detection of corrupted or duplicate data
- **Data labelling process**: number of labels, type of label annotation (human and/or AI), type of label (bounding box, segmentation, text), number of human annotators, setup for the labelling process (software platform), cross-checks for mislabeled data…
- **Splitting of the dataset**: rationale for the data splitting strategy, number and/or percentage of images in training, validation and test dataset; distribution of features within the split dataset.
- **Data leakage**: Strategy to avoid leakage of data from training to the test dataset. Clear indication if the data splitting was performed before or after any image processing.
- **Image processing**: raw data used, data dimensionality reduction (cropping, conversion to black and white), formatting, naming conventions, normalization of data
- **Statistical characterization of the dataset**: label distribution (counts of labels in the dataset), over- or underrepresentation of classes (for example rare defects)

These are examples of elements that support form a comprehensive dataset characterization framework that helps ensure the quality and integrity of data used in AI/ML training. However, no

specific requirements or guidance to systematically describe large datasets used for training AI/ML models in pharmaceutical manufacturing are mandatory yet. To bridge this gap, principles and elements of the "Assessment List for Trustworthy Artificial Intelligence (ALTAI)" (123) by the independent high-level expert group on AI of the European Commission, as well as data science best practices from the growing research field related to bias in AI (124–127), may be used to demonstrate transparency, traceability, explainability, and trustworthiness.

The importance of these data handling practices should be also put in perspective, since AI/ML model training may require thousands to millions of datapoints in order to generate a model that is fit for use.

### 4.2.3 Change Management

Once the initial validation activities are finished and the AI/ML system or sub-system is frozen (no more training) and deployed, the model is given a version number and put under change control (92). Operational controls and manual verification of results should be defined in a predetermined monitoring plan with limits to identify critical performance drifts that indicate that retraining is necessary. The performance must be monitored because data properties and statistical distribution of features may change over time (data drift: initial dataset quality VS dataset quality in operation) (128,129). In the case of AVI, scenarios such as identification of new defect types, changes in primary packaging components or drop in performance metrics such as increased FRR can indicate a change in the distribution of the input-data in operation and may trigger retraining. Renato Panza presented a risk-based approach to change control where, depending on affected component of the AI/ML system (training data, code, hyperparameters,…) and the level of risk the change is introducing, activities could either fall under “formal change” or under “regular maintenance activities” (93):

- Major versioning is triggered by formal changes. For example, changes in code or hyperparameters or new dataset sources.
- Minor versioning is triggered by regular maintenance activities. For example: retraining with supplemental data.

However, AI/ML specific risks such as “catastrophic forgetting” need to be taken into account and evaluated, where previously learned information can abruptly be forgotten upon retraining (130). Hence, change control procedures and versioning of datasets used for training and testing may be necessary to establish traceability and transparency in order to reconstruct how the model has been trained, similar to the use of laboratory notebooks to reconstruct experiments. In analogy to “data cards”, the concept of “model cards” was proposed to standardize documentation of changes to the model (131). Refer to section **4.4.3.2** for an example from the medical device regulatory framework.

### 4.2.4 Summary (RQ2)

The literature review found that a risk-based approach is critical for ensuring both patient safety and product quality in the integration of AI/ML systems within pharmaceutical manufacturing. Key findings indicate that established frameworks - such as ICH Q9, the EU AI Act, and the US NIST AI RMF - provide a robust basis for categorizing and mitigating risks. A holistic risk management framework, showed by the ISPE ML Risk and Control Framework, effectively guides the entire lifecycle from risk assessment through control and review, while also adapting computerized system validation processes to address AI-specific challenges. Moreover, this review highlights the fundamental role of robust data governance and change management practices to ensure data integrity and traceability, as well as the importance of explainable AI (xAI) to overcome the “black-box” nature of complex DL models. External risks related to supplier reliability, IT infrastructure, and cybersecurity further underscore the need for a comprehensive approach to risk management.

With these operational and risk management considerations in mind, it is equally important to understand the evolving legal frameworks that govern AI/ML applications in GMP environments (RQ3).

### 4.3 AI/ML in GMP: Legal Framework in Europe and the United States of America (RQ3)

The European Council unanimously approved the world's first comprehensive and industry agnostic Artificial Intelligence regulation 2024/1689 or "AI Act" on the 21st of May 2024 (132). This regulation does apply to the pharmaceutical industry, however, at time of writing, neither AI nor ML are mentioned in binding sector specific regulations or GMP guidelines (86). The European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) are aware of the rapidly evolving need for guidance and both issued discussion papers with requests for industry feedback, see **TABLE 20**.

| Agency | Document title | Document type | Reference | Publishing date |
|---|---|---|---|---|
| EMA | Concept Paper on the revision of Annex 11 of the guidelines on Good Manufacturing Practice for medicinal products – Computerised Systems | Concept paper | (133) | 19.09.2022 |
| EMA | Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle | Draft Reflection paper | (134) | 13.07.2023 |
| EMA | Multi-annual artificial intelligence workplan 2023-2028: HMA-EMA joint Big Data Steering Group | Workplan | (135) | 18.12.2023 |
| EMA | Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle | Final Reflection paper | (52) | 09.09.2024 |
| EMA & PIC/S | Draft guidelines: New annex 22 – Artificial intelligence | Draft guidance | (136) | 07.07.2025 |
| FDA | Artificial Intelligence in Drug Manufacturing | Discussion paper | (94) | 01.03.2023 |
| FDA | Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products | Discussion paper | (137) | 10.05.2023 |
| FDA | Artificial Intelligence & Medicinal Products: How CBER, CDER, CDRH and OCP are Working Together | Area of Focus paper | (138) | 15.03.2024 |
| FDA | Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products | Draft guidance | (139) | 06.01.2025 |
| FDA | Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products | Discussion paper | (137) | 10.05.2023 |

*Table 20 Overview of key EMA and FDA documents related to AI in pharmaceutical manufacturing.*

The high-level approaches of regulating AI in the EU and USA (140) are compared in **TABLE 21**.

| Level of Regulation | European Union (EU) | United States of America (USA) |
|---|---|---|
| Legally Binding Acts (**Hard Law**) | - EU AI Act<br>- GDPR | - No overarching federal AI law<br>- Sector-specific laws |
| National / Member State Laws | - Transposition of EU directives<br>- Additional member-state legislation | - Federal laws<br>- State-level laws |
| Policies, Strategies & Principles | - EU Coordinated Plan on AI<br>- High-Level Expert Group on AI Ethics Guidelines | - National AI Strategy<br>- Blueprint for an AI Bill of Rights (141) |
| Non-Binding Frameworks (**Soft Law**) | - Ethics Guidelines for Trustworthy AI | - NIST AI Risk Management Framework (83) |

*Table 21 Comparative overview of AI regulatory levels and frameworks in the EU and the US.*

Research question 3 (RQ3) will discuss the EU and United States GMP-specific evolving legal framework and feedback from pharmaceutical industry associations.

#### 4.3.1 Regulatory initiatives by EMA

In Europe, European Medicines Agency (EMA) and the Heads of Medicines Agencies (HMA) have published the "Multi-annual AI workplan 2023-2028" under the joint "HMA-EMA Big Data Steering Group" to respond to the new opportunities and challenges of AI (135). The EMA Quality Innovation Group (QIG) was established in September 2022 as an expert group to support innovative approaches, such as AI, for the development, manufacture, and quality control of medicines (142). The current version of EU GMP "Annex 11: Computerised Systems" has been in force since 2011 and applies to all

computerized systems used as part of GMP regulated activities but lacks guidance on AI among other emerging technologies as pointed out in the concept paper on the revision of Annex 11 (14,133):

*"There is an urgent need for regulatory guidance and expectations to the use of artificial intelligence (AI) and machine learning (ML) models in critical GMP applications as industry is already implementing this technology. The primary focus should be on the relevance, adequacy and integrity of the data used to test these models with, and on the results (metrics) from such testing, rather that on the process of selecting, training and optimising the models."*

The concept paper on the revision of Annex 11 was published in September 2022 and includes a timetable with a proposed release date of the draft guideline for December 2024 followed by adoption by EMA and PIC/S in 2026 (133). As part of the HMA-EMA joint Big Data Stering Group workplan, a draft scientific guideline "Reflection paper on the use of AI in the medicinal product lifecycle" was published on 19.07.2023 by EMA (134) with an open public consultation window for stakeholder feedback until 21.11.2023. EMA received 1342 comments from 66 stakeholders including regulatory bodies, public and private consortia, organizations and individuals (143). The updated final reflection paper was published on the 9th of September 2024 and includes the definition of "AI system" developed by the Organisation for Economic Co-operation and Development (OECD) as well as replacing the term "high risk", which is used within the "AI Act" (15) (see **FIGURE 13**) for risk classification, with "high patient risk" and "high regulatory risk" (143). Technical aspects such as:

- data acquisition and augmentation,
- training, validation, and test datasets,
- model development,
- performance assessment,
- interpretability and explainability and
- model deployment

as well as governance, data integrity / protection and ethical aspects of trustworthy AI are covered by the scientific guideline. However, in contrast to the EU legislative framework (directives, regulations, decisions…), scientific guidelines are intended to promote dialogue with stakeholders and are not legally binding but rather express a viewpoint. Therefore, alternative approaches by pharmaceutical manufacturers are acceptable with proper justification. Note that reflection papers express viewpoints on a topic, regulation as well as use and are intended to promote dialogue with stakeholders.

#### 4.3.1.1 Draft guideline: New EU GMP Annex 22 – Artificial Intelligence

The European Commission opened stakeholder's consultation on 07.07.2025 with a deadline for public health stakeholders involved in GMP activities ending on 07.10.2025. Three guidelines were drafted by the EMA GMDP-Inspectors Working Group in cooperation with the PIC/S with the aim to support innovation in manufacturing of medicines and ensuring regulatory harmonization, see **TABLE 22**.

| Document | In Operation (Date / Pages) | Draft (Date / Pages) |
|---|---|---|
| Chapter 4 – Documentation | 30.06.2011 / 9 pages | 07.07.2025 / 17 pages |
| Annex 11 – Computerised Systems | 30.06.2011 / 5 pages | 07.07.2025 / 19 pages |
| Annex 22 – Artificial Intelligence | - | 07.07.2025 / 6 pages |

*Table 22 Overview of current and draft EU GMP documents. The table compares the operational and draft status of Chapter 4 (Documentation) and Annex 11 (Computerised Systems) as well as the draft of the new Annex 22 (Artificial Intelligence), including their respective publication dates and page counts.*

The simultaneous revision of Chapter 4 (144), Annex 11 (145) and the introduction of a new Annex 22 (136) signals an alignment towards a digitally mature regulatory framework for GMP. Chapter 4 on documentation establishes a risk-based data governance system within the pharmaceutical PQS as the foundation for data integrity. Any data generated or used by AI models, for example training data, must

meet standards for GMP records (ALCOA++). The revision of Annex 11 covers the infrastructure that AI models, as a software sub-system, uses. It clearly outlines supplier and service provider oversight, including XaaS arrangements (see **FIGURE 17** on different XaaS models), to emphasize that manufacturers retain full regulatory responsibility even when AI capabilities are outsourced. Annex 22 builds on the CS-framework of Annex 11 by adding guidance on AI used in critical applications that directly affect patient safety, product quality, or data integrity. The significance of AI in manufacturing is highlighted by the introduction of this new Annex, rather than incorporating AI topics into the revision of Annex 11. This approach enhances regulatory adaptability by allowing more frequent revision cycles the new Annex 22 in the rapidly changing AI-field while preserving the stability of more permanent rules for CS in Annex 11.

Notably, Annex 22 defines the scope of how AI/ML can be used in manufacturing of medicinal products and active substances. Only static models that do not adapt their own performance during use (see **TABLE 13** on AI Maturity) are allowed for critical GMP applications. Generative AI (GenAI) that creates new data rather than analyzing or classifying existing data and Large Language Models (LLM) are excluded from the use in critical GMP applications. For non-critical GMP applications (no direct impact on patient safety, product quality or data integrity), the new Annex 22 leaves the door open for novel AI tools if personnel with adequate qualification and training ("human-in-the-loop") ensures that the output of such models are suitable for the intended use. The document itself is structured in ten sections and, like the revision of Chapter 4 and Annex 11, appends a glossary (see **TABLE 23**).

| Section | Key takeaways |
|---|---|
| 1. Scope | Defines the scope for AI/ML use in critical GMP applications. Only static, deterministic models are allowed. Excludes adaptive models, LLMs, and GenAI for critical use. Human-in-the-loop may be required for non-critical applications. |
| 2. Principles | Emphasizes QRM, roles and responsibilities, qualified personnel that understand the intended use and the associated risks of an AI model in a GMP environment, and robust documentation. Cross-functional collaboration (SME, QA, data scientists, IT and consultants) is expected. |
| 3. Intended Use | Clearly define the intended use of the model, limitations, and subgroups of the input sample space. Description of operator responsibility in human-in-the-loop as well as training and performance monitoring for such operators. |
| 4. Acceptance Criteria | Test metrics and acceptance criteria (e.g., accuracy, F1-score). Must ensure that process performance is not degraded (e.g., via relevant KPIs). |
| 5. Test Data | Data selection, dataset size, labeling approach, and data pre-processing steps. Ensure exclusion of biased or reused data. Synthetic test data or labels is not recommended, and any use must be fully justified. |
| 6. Test Data Independency | Ensure independence between training, validation, and test datasets. Prevent data leakage through staff separation, handling of data from physical objects (for example images of the same object), and identification of data usage (which data, when, how many times). |
| 7. Test Execution | Tests must ensure that the model is fit for the intended use the model should generalize well. Include documented test plans, approvals, possible deviations, and test result traceability. Process SMEs should be involved in development of the test plan. |
| 8. Explainability | Feature attribution tools (for example SHAP or LIME) or visual tools (heat map) to highlight key factors contributing to the outcome. Ensure that the model makes decisions based on relevant features (feature justification). |
| 9. Confidence | Models must produce a confidence score and flag uncertain predictions. Thresholds must be defined to suppress or escalate unreliable results for review. |
| 10. Operation | Covers change control, system performance monitoring, input sample space monitoring, configuration control, and ensuring ongoing human review / oversight (human-in-the-loop). |
| Glossary | AI, DL, feature, LIME, ML, model, overfitting, SHAP, static, test- & training dataset, validation data (in AI) |

*Table 23 Summary of the structure and key takeaways of the 07.07.2025 draft of Annex 22 - Artificial Intelligence. Each section of the document is briefly explained, highlighting its main focus areas.*

### 4.3.2 Regulatory initiatives by FDA

In the United States, the Food and Drug Administration (FDA) regulates medicinal products through the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) to ensure that standards for quality, safety, and efficacy are met. Innovative approaches to manufacturing and the corresponding technical and regulatory challenges for pharmaceutical companies are addressed by CDERs Emerging Technology Programme (ETP) and CBERs Advanced Technologies Team (CATT) to facilitate collaborative discussion, identifying and resolving of potential

technical and regulatory issues regarding the development and implementation (146). In 2019 the Artificial Intelligence Steering Committee (AISC) was established by CDER to provide guidance, oversight and to come up with a strategy on the integration of AI in drug development, see **FIGURE 22**. In 2021 CDER initiated the Framework for Regulatory Advanced Manufacturing Evaluation (FRAME) to support the adoption of advanced manufacturing technologies from a policy point of view with one of the priority areas being the application of AI. Leveraging and continuing this initiative is part of the 2024 collaborative area of focus of FDAs medical product centers to "Promote the Development of Standards, Guidelines, Best Practices, and Tools for the Medical Product Life Cycle" (138). The document describes four AI focus areas and how CDER, CBER, CDR and the Office of Combination Products (OCP) are jointly working together to cover the whole lifecycle of medicinal products.

In a response to Executive Order 14110 from October 2023, CDER AISC as well as the AI policy working group and AI Community of Practice were merged into the "CDER AI Council" in 2024. On 20th of January 2025, the Executive Order was rescinded by the subsequent administration with, at the time of writing, unclear implications on CDERs AI initiatives. Shortly before, on the 7th of January 2025, the draft guideline "Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products" was published by FDA (147). This draft guidance is being issued consistent with FDAs good guidance practices regulation (21 CFR 10.115): First, external input via public meetings and workshops is obtained. Second, the draft guidance is published with an invitation to public comments (current stage at the time of writing). Finally, the feedback will be reviewed, incorporated, and the guidance will be finalized.

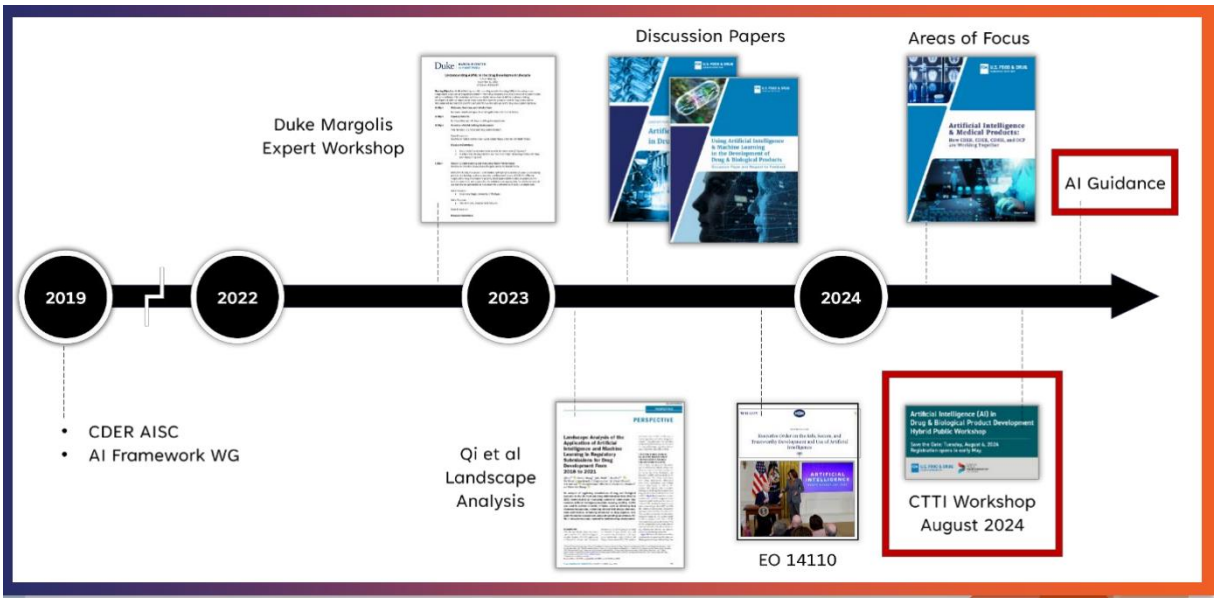


*Figure 22 CDER AI Policy Milestones slide from the presentation "Responsive Regulation of Artificial Intelligence in Drug Develoment" at the Regulatory Education for Industry (REdI) Conference 2024 by Tala H Fakhouri, Associate Director for Policy Analysis; CDER | US FDA (148).*

The development of a regulatory framework was approached by issuing two discussion papers, one focusing on drug manufacturing, and conducting workshops to gather feedback and insights from various stakeholders outside the FDA (pre-draft input). This process resulted in the creation of an FDA center-wide AI draft guidance document in January 2025 that is open for comments and suggestions for 90 days until 07.04.2025. The draft guideline introduces a seven-step risk-based credibility assessment framework that is used for establishing and evaluating the credibility of AI-models for a particular "context of use" (COU) (139).

The draft guidance describes an example for commercial manufacturing of a drug that is a parenteral injectable dispensed in a multidose vial, with the filled volume of preparation being a critical quality attribute for the release of the drug. In the example, the manufacturer is proposing to implement an AI-based visual analysis system to perform 100% automated assessment of the fill level in the vials of Drug B to enhance the performance of the visual analysis system and to identify deviations, see **TABLE 24**.

| Step | Example for commercial manufacturing |
|---|---|
| 1. Define the Question of Interest | Do vials of Drug B meet established fill volume specifications? |
| 2. Define the COU for the model | An AI-based model will be used to analyze data obtained from visual images of the vials to determine if a deviation in volume has occurred (the AI model's role). However, as part of the release testing, independent verification of the fill volume is performed on a representative sample for each batch. Therefore, the AI-based model will *not* be the sole determinant for the release of product (the AI model's scope). |
| 3. Assess the AI model risk | Deviations in the volume of vials containing Drug B could result in a number of issues. For example, the release of units that do not meet quality standards could potentially lead to medication errors due to either an inability to withdraw labeled content or pooling of vials to obtain a single dose (if not identified in labeling). Because volume is a critical quality attribute and incorrect volume measurements would have a high impact on product quality, the decision consequence would be high. However, for this example, a manufacturer, as a part of release testing, would measure fill volume on a representative sample for each batch. Measuring fill volume through release testing would reduce the AI model influence, and therefore the model influence would be determined to be low. Given that the decision consequence is deemed high and the model influence is deemed low with the stated mitigations, the model risk for this COU is medium. |
| 4. Develop a plan to establish the credibility of AI model output within the COU | The credibility assessment activities are intended to provide a general list of activities that should be considered when establishing the credibility of AI model outputs. The appropriate credibility assessment activities may vary depending on the nuances of a specific development program that cannot be captured in the hypothetical examples provided. |
| 5. Execute the plan | Step 5 to 7 Relate to step 4, as they are intended recommendations to execute, document, and assess the credibility assessment activities of step 4. |
| 6. Document the results of the credibility assessment plan and discuss deviations from the plan | |
| 7. Determine the adequacy of the AI model for the COU | |

*Table 24 Proposed seven-step process for AI model implementation in commercial manufacturing - example (139)*

FDAs guidance document outlines the Agency's views on regulatory issues. They are not legally binding and hence do not establish legally enforceable responsibilities but rather provide recommendations. Alternative methods can be used if they meet statutory and regulatory requirements.

### 4.3.3 Stakeholder Feedback to Regulatory Initiatives

Both EMA and FDA received feedback from stakeholders such as associations, academia, private companies and individuals. The publicly accessible feedback was accessed via FDAs comment section for the dockets of the two discussion papers released in 2023. Comments on "Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle" received by EMA are not publicly available and hence no further analysis of stakeholder feedback was possible.

#### 4.3.3.1 Feedback to FDA

As per FDA, 800 comments were received in total. Among these, 68 public comments are accessible for the discussion paper "Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products" (Docket FDA-2023-N-0743) and 49 public comments are available for the discussion paper "Artificial Intelligence in Drug Manufacturing" (Docket FDA-2023-N-0487). Among the 49 public comments for AI in drug manufacturing, 43% included visual inspection in their answer to the FDA request for feedback on "What types of AI applications do you envision being used in pharmaceutical manufacturing?". This data shows that pharmaceutical industry is clearly interested in this particular application of AI within manufacturing.

According to a presentation of Dr. Tala Fakhouri, Associate Director of Policy Analysis, at the Regulatory Education for Industry (REdI) Conference 2024, the responses to the discussion papers can be summarized largely into the following topics (107):

1. Clarity on what falls within/outside the scope of FDA's oversight
2. Clarity on how to operationalize a risk-based approach
3. Clarity on "transparency" and the level of detail and documentation required
4. Calls for harmonization globally, and alignment with medical devices
5. Calls for establishment of partnerships to advance the creation/sharing of machine-readable data sets for drug development

The stakeholder feedback resulted in a 20-page draft guidance that, apart from the topic of data sharing, provides recommendations to sponsors and other interested parties (depending on the stage of the drug product life cycle) on the use of AI *"…to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality of drugs"*.

#### 4.3.3.2 Feedback to EMA

The CHMP and CVMP adoption of the draft "Reflection paper on the use of artificial intelligence in the medicinal product lifecycle" was followed by one-year public consultation that ended on the 31st of December 2023 and EMA received a total of 1342 comments from 66 stakeholders (143). EMA published a document about the implementation of the comments and summarized the most important changes to the final document on a general note (143):

1. Technical terms and definitions have been harmonized and are now supported by an expanded glossary
2. Clarity on the scope of the reflection paper to not exclude shallow ML technology
3. Clarity on terminology in relation to risk has been implemented to create a clear separation between the term "high risk AI" in Annex 3 of the AI act. The term has been replaced with "high regulatory impact" and "high patient risk" to clarify that most regulatory requirements are founded in medicines legislation
4. Clarity regarding requirements on availability of data used in modelling, use of third-party AI systems, and need for prospective testing of machine learning models

EMA further explained that many detailed comments that did not make it to the reflection paper but will instead be taken into consideration for the future drafting process of formal EMA scientific guidelines in the field of AI.

The public consultation on the draft of the new Annex 22 - Artificial Intelligence is ongoing with a deadline for feedback on 07.10.2025.

### 4.3.4 Initiatives by National Competent Authorities

In response to the foreseeable integration of AI in GMP processes, few European agencies have launched initiatives to prepare for the gap between traditional static computerized systems and the dynamic nature of AI/ML applications.

#### 4.3.4.1 Denmark

Notably, the Danish Medicines Agency Inspector Ib Alstrup published a two-page draft document called "Questions to critical GxP AI/ML applications" in February 2021. The document is composed of 16 questions that are sectioned into the following topics (58):

- Datasets
- Bias and Variance

- Confusion Matrix and Metrics
- Interpretation of Results

This document is intended to facilitate the critical assessment of GxP AI/ML applications that utilize static algorithms and supervised learning techniques. It provides comprehensive questions related to data management, algorithm training, performance evaluation, and result interpretation.

In August 2021, BioPhorum, an organization representing the interests of the biopharmaceutical industry, was among the first to publicly submit feedback in response to the questions posed by DMKA including several requests for topics to include in future guidelines (149).

#### 4.3.4.2 Germany

In Germany, the central authority for the health protection with regard to medicinal products and medical devices "Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten" (ZLG) published an Aide-Mémoire on inspection of computerized systems that is used by German inspectors in preparation for inspections. The document states (translated from German to English) the following about validation of systems that use AI in section 4. Validation (110):

*"Computerized systems that employ artificial intelligence—that is, systems that make decisions not based on predefined algorithms but rather on training data through complex, evolving algorithms—cannot be validated using this classical approach. To date, a unified European standard for the requirements of validation is not available. The primary focus should be on the relevance, appropriateness, and integrity of the data used for testing these models, as well as on the outcomes of these tests and, where available, in comparison with already accepted standards."*

The Aide-Mémoire, while not directly mentioned, makes it evident that inspectorates see a gap in requirements for computerized system validation with AI components.

### 4.3.5 Summary (RQ3)

The review of regulatory initiatives reveal that both the EMA and FDA are actively shaping regulatory frameworks to accommodate AI/ML in GMP environments. In Europe, the AI Act and the issuance of concept and reflection papers by the EMA underscore an urgent need to update existing guidelines—such as EU GMP Annex 11—to address AI-specific challenges. The European Commission addressed this need in July 2025 by publishing a new draft guidance document: Annex 22, specifically focused on the use of AI in GMP-regulated activities. In parallel, revised drafts of Chapter 4 (Documentation) and Annex 11 (Computerised Systems) were also released, aiming to strengthen data governance and digital system oversight as foundational elements for AI integration. In the United States, the FDA has similarly advanced its regulatory approach by releasing discussion papers that resulted in a draft guidance that introduces a seven-step risk-based credibility assessment framework for AI in drug manufacturing. Additionally, national competent authorities in Denmark and Germany have released supplementary guidance documents for GMP inspections to bridge the gap between requirements for traditional rule-based and AI/ML CS applications. Collectively, these efforts reflect a converging international focus on ensuring that AI/ML integration within pharmaceutical manufacturing is both safe and compliant.

Insights from these regulatory initiatives are matched by best practices emerging from other highly regulated industries, which offer valuable guidance for further refining AI/ML applications in pharmaceuticals (RQ4).

### 4.4 AI/ML in the medical device industry: Best practices (RQ4)

This section focuses on the medical device industry, which, much like the pharmaceutical sector, operates under stringent regulatory frameworks such as US FDA guidelines, EU medical device regulation (MDR) as well as in-vitro Diagnostic regulation (IVDR), and ISO 13485.

While other highly regulated industries - such as automotive, nuclear, and aviation - also implement advanced AI/ML systems, the parallels between medical devices and pharmaceuticals make the former particularly relevant for gaining insights. These best practices may offer valuable approaches for the use of AI/ML (sub)-systems in pharmaceutical manufacturing in a GMP setting.

The medical device industry needed to rapidly establish regulatory AI/ML frameworks to catch up with the exponential growth of AI-adoption. Between the year 2010 and 2023, 836 AI/ML-enabled medical devices were authorized for marketing in the United States alone, see **FIGURE 23**. Among all these devices, 649 (78%) were authorized for applications in the specialty field of radiology, see **FIGURE 24**. Radiology and pharmaceutical visual inspection AI/ML applications are similar in many ways, as both use advanced image processing techniques to detect, locate and quantify anomalies.

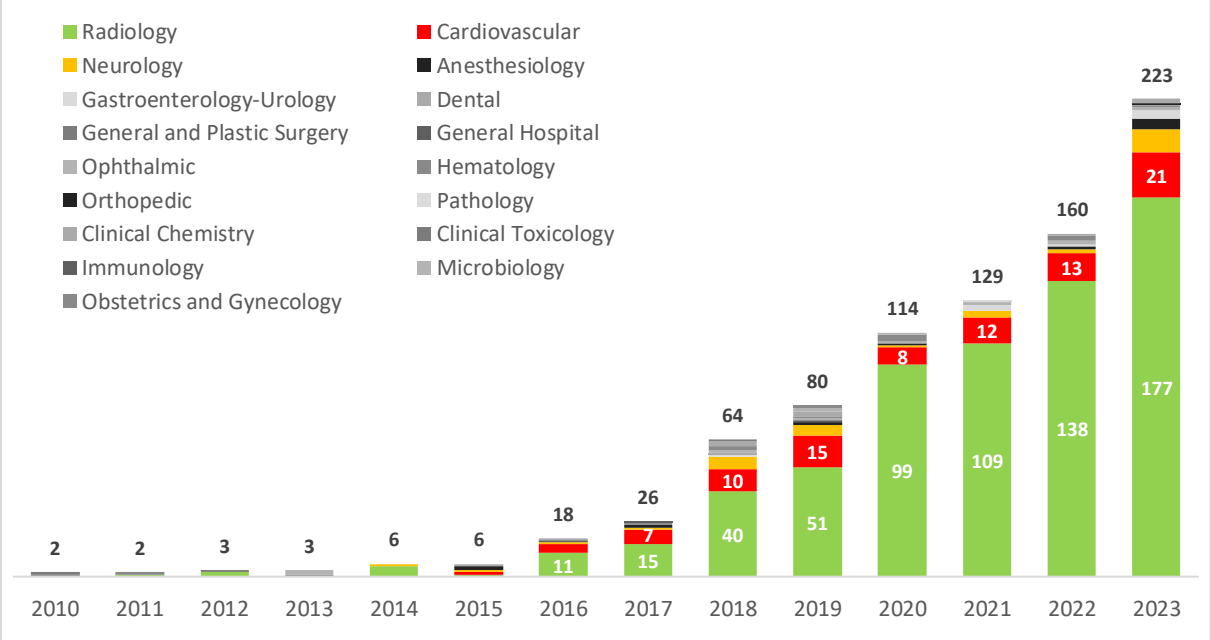


*Figure 23 Number of AI/ML-enabled Medical Devices authorized for marketing in the United States from 2010 – 2023. Data is based on the list of "AI/ML-enabled Medical Devices" available on the FDA website (150).*

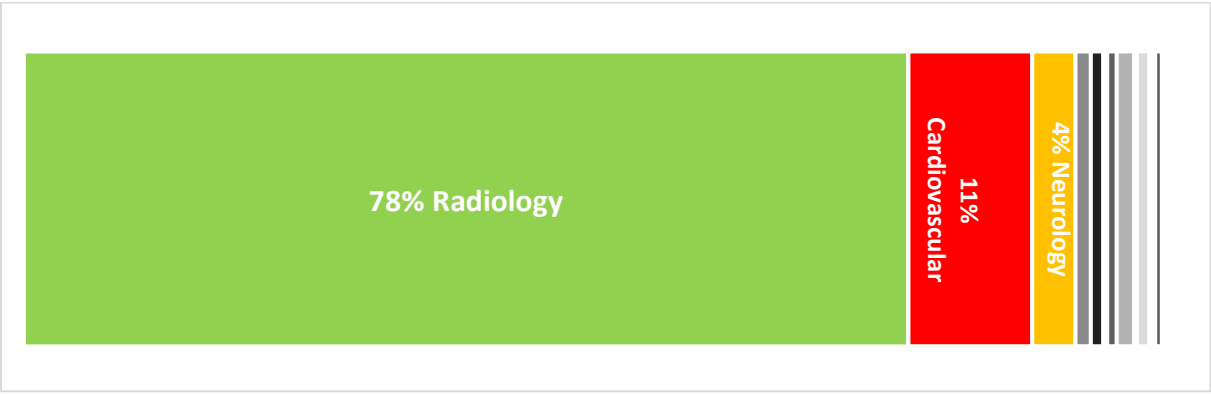


*Figure 24 Distribution of FDA-assigned medical specialty field for AI/ML-enabled Medical Devices between 2010 and 2023 with radiology (78%), cardiovascular (11%) and neurology (4%) having the highest representations among all devices (n=836).*

FDA has recently shifted towards a more collaborative regulatory approach for AI/ML-enabled medical products. Initially, the Center for Devices and Radiological Health (CDRH) independently released guidance documents addressing AI/ML in medical devices (151,152). However, over the past two years, there has been a marked transition towards cross-center collaboration. This is noticeable by the March 2024 release of the common document "*Artificial Intelligence & Medicinal Products: How CBER, CDER, CDRH and OCP are Working Together*", which underscores the integration of expertise across multiple product centers. In addition, the development of common guidance documents - such as (153), (139) and (154) - further illustrates the FDA's commitment to a harmonized regulatory framework.

Notably, the highly anticipated draft guidance "*Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products*" adopts high-level key concepts and principles of the risk-based credibility assessment framework from the FDA-recognized consensus standard for medical devices titled "*American Society of Mechanical Engineers Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*" (ASME V&V 40) (139).

Among the unique challenges that AI use presents, FDA lists the following in the draft guideline:

1. Data and data management: Bias and reliability of AI-driven results due variability in the quality, size and representativeness of datasets for training AI models. Hence, data used to develop AI models should be fit for use (=relevant and reliable).
2. Methodological transparency: Due to the complexity it is necessary to understand how AI-models have been developed and how they arrive at their conclusions.
3. Accuracy uncertainty of the model output: This aspect is needed to interpret, explain and quantify the deployed model output.
4. Change management: Due to data drift, the performance of the deployed model might change over time if it receives input that is different compared to the training data.

The following sub-chapters discuss additional medical device guidelines and guiding principles that can be used to address these challenges in addition to the recommendations in (139).

### 4.4.1 American Society of Mechanical Engineers (ASME)

The ASME V&V 40 was published in 2018 by the American Society of Mechanical Engineers (ASME) to address the lack of harmonized standards to assess the credibility of computational models. The ASME subcommittees develop discipline specific standards, such as *V&V 30 – Verification and Validation in Computational Simulation of Nuclear System Thermal Fluids Behaviour*, to help frame the credibility expectations of computational models and provide pathways to regulatory adoption.

The core principles of the framework to ensure credibility of computational models are (155):

- "Question of Interest",
- "Context of use" (COU),
- Assessment of model risk.

While ASME V&V 40 has been specifically developed for physics-based models for medical devices, these core principles should, according to FDA, ensure credibility for the use of AI models to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs.

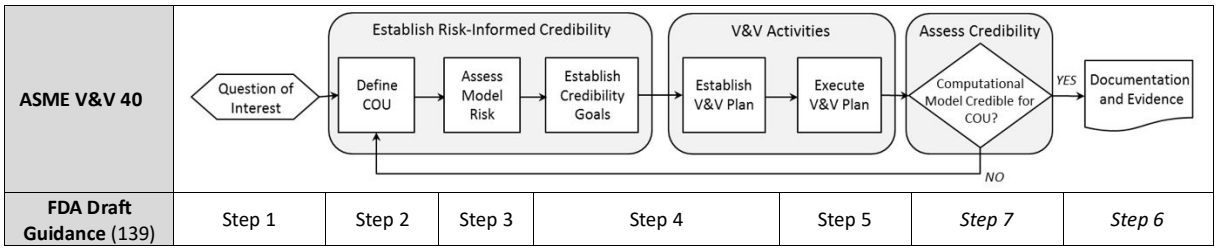


*Figure 25 Comparison between ASME V&V40 risk assessment framework and the 7-step process outlined in the FDA draft guidance (139)*

**TABLE 24** presents an example of the 7-step process. The model risk depends on two main factors related to the question of interest:

1. Model influence: "…contribution of the evidence derived from the AI model relative to other contributing evidence used to inform the question of interest."
2. Decision consequence: "…describes the significance of an adverse outcome resulting from an incorrect decision concerning the question of interest."

The resulting *model risk* is important since it is used to modulate extent of the credibility assessment activities to ensure appropriate credibility of the AI-model outputs. Parallels from this risk assessment approach can be drawn to the ISPE D/A/CH Working group proposed *AI Maturity Model for GxP Applications* which use similar control design called "AI system maturity" and "Risk impact", see Erdmann N et al (87) or section **4.2.1 RISK MANAGEMENT**.

### 4.4.2 Joint FDA, Health Canada and MHRA Guiding Principles for Medical Devices

Since 2021 the US FDA, Health Canada and Medicines and Healthcare products Regulatory Agency (MHRA) have jointly published a series of guidance documents to specifically address AI/ML-enabled medical devices, see **TABLE 25**.

| Document title | Document type | Reference | Publishing date |
|---|---|---|---|
| Good Machine Learning Practice for Medical Device Development | Guiding Principles | (156) | October 2021 |
| Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices | | (157) | October 2023 |
| Transparency for Machine Learning-Enabled Medical Devices | | (158) | June 2024 |

*Table 25 Overview of key guiding principles for ML-enabled medical devices*

Good Machine Learning Practices (GMLP) are intended to promote safe, effective, and high-quality medical devices that use AI/ML. One of the greatest potential benefits of AI/ML is the ability to improve model performance through iterative modifications, including learning from real-world data. To support this intrinsic iterative nature of AI/ML, the concept of a Predetermined Change Control Plans (PCCP) was introduced (157) as an extension of principle 10 of GMLP (156) to facilitate the change management process; the transparency guiding principles (158) can be seen as an extension of GMLP principle 7 and 9. Due to their universal nature, these principles may be also used by other healthcare sectors such as the pharmaceutical medicinal products sector. **TABLE 26** summarizes the guiding principles.

| **Good Machine Learning Practice for Medical Device Development: Guiding Principles** (156) | |
|---|---|
| 1. | Multi-Disciplinary Expertise Is Leveraged Throughout the Total Product Life Cycle |
| 2. | Good Software Engineering and Security Practices Are Implemented |
| 3. | Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population |
| 4. | Training Data Sets Are Independent of Test Sets |
| 5. | Selected Reference Datasets Are Based Upon Best Available Methods |
| 6. | Model Design Is Tailored to the Available Data and Reflects the Intended Use of the Device |
| 7. | Focus Is Placed on the Performance of the Human-AI Team |
| 8. | Testing Demonstrates Device Performance During Clinically Relevant Conditions |
| 9. | Users Are Provided Clear, Essential Information |
| 10. | Deployed Models Are Monitored for Performance and Re-training Risks Are Managed |
| **Predetermined Change Control Plans for Machine Learning: Guiding Principles** (157) | |

| | |
|---|---|
| 1. | Focused and Bounded: A PCCP describes specific changes that a manufacturer intends to implement. Such changes are limited to modifications within the intended use or intended purpose. |
| 2. | Risk-based |
| 3. | Evidence-Based |
| 4. | Transparent |
| 5. | Total Product Lifecycle (TPLC) Perspective |
| **Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles** (158) | |
| 1. | Who: Relevant audiences for transparency |
| 2. | Why: Motivation for transparency |
| 3. | What: Relevant information |
| 4. | Where: Placement of information |
| 5. | When: Timing of communication |
| 6. | How: Methods to support transparency |

*Table 26 Summary of guiding principles for ML-enabled medical devices in the documents listed in Table 25*

An overview of regulatory and industry initiatives and GMLP in GxP regulated areas was discussed by Stueben J et al (159). The authors conclude that the US is showing a more progressive approach in enabling innovation compared to Europe and recommend an overarching harmonization of regulatory expectations to foster adoption of AI/ML methodologies. The authors further urge that regulators promptly define requirements to secure ongoing and planned developments and investments of the pharmaceutical industry.

### 4.4.3 FDA Medical Device Guidance Documents

The joint guiding principles are reflected in the recently published FDA draft guidance and guidance documents that also contain best practices that may be relevant to the use of AI/ML in GxP settings.

| Document title | Document type | Reference | Publishing date |
|---|---|---|---|
| Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions - Guidance for Industry and Food and Drug Administration Staff | Draft Guidance | (160) | 03.04.2023 |
| | Guidance | (153) | 04.12.2024 |
| Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations – Draft Guidance for Industry and Food and Drug Administration Staff | Draft Guidance | (154) | 07.01.2025 |

*Table 27 Reviewed FDA medical device guidance documents.*

#### 4.4.3.1 Predetermined Change Control Plan (PCCP)

Medical device manufacturers can include a Predetermined Change Control Plans (PCCP) in FDA marketing submissions for medical devices that includes AI-enabled device software functions (AI-DSF) (153). The FDA reviews the PCCP in a marketing submission to ensure device safety and effectiveness, without needing extra submissions for each modification that is described within the version-controlled PCCP. The PCCP consists of three main sections:

1. Description of Modifications
2. Modification Protocol
3. Impact Assessment

The “Modification Procotol” is at the heart of the PCCP and includes four primary components. Example elements of each component are provided in the Appendix A of the guidance document. These examples are guided by questions that may be relevant for AI/ML applications in GxP settings since all four components are encountered in the AI/ML-software lifecycle. For example, Appendix A includes the following subsections with further detailed questions:

1. Data Management
    a. Collection protocols
    b. Assurance of data quality
    c. Reference standard determination
    d. Sequestration of test data sets

2. Re-Training
    a. Re-Training objectives and focus
    b. Re-Training implementation
3. Performance Evaluation
    a. Triggers to initiate performance evaluation
    b. Assessment metrics and elements
    c. Statistical analysis plans
    d. Performance targets
    e. Additional testing needs
4. Update Procedures
    a. Software verification and validation
    b. When and how updates will be implemented
    c. Communication and transparency to users
    d. Device monitoring plan

The section on data management is particularly interesting since topics and questions may be also relevant to GMP inspections to understand the manufacturers data management practices. For example, some questions of the guideline are paraphrased below (153):

- How does the manufacturer plan to obtain and use training, tuning, and test data that are complete and representative of the proposed intended use and intended environment?
- Are identifiable subpopulations adequately represented and separated into training, tuning, and test sets to minimize AI model bias?
- How will training, tuning, and test data be sequestered to prevent overfitting and misquotes of test performance?
- How will older data be complemented or replaced by newer data so that the performance is representative (of the current patient population and standard of care)?
- Whether the reference standard represent the best available process for determining the ground truth?
- How will new data be collected, annotated, curated, stored, retained, controlled, and used?

Refer to RQ2 section **4.2.2 DATA GOVERNANCE** on challenges related to data management topics. The concept of PCCPs is a forward-thinking approach that facilitates the iterative nature of AI/ML software while at the same time promotes safe and effective AI-enabled devices. FDAs recent draft guidance for pharmaceutical drug and biological products (139) is heading into a similar direction and describes a "postapproval change management plan" that prospectively manages model-related changes to "established conditions" over the drug product life cycle which would not require a submission to the agency prior to making modifications.

#### 4.4.3.2 AI-DSF Lifecycle Management

Potentially relevant topics included in the FDA draft guidance document (154) are the sections about performance monitoring, cybersecurity and an appendix about the use of "model cards". The draft guidance recognizes that AI/ML models are highly dependent on the characteristics of data used to train them and hence their performance can be particularly sensitive to changes in data input (data drift) (154). If the AI/ML-enabled software is part of a highly automated process with limited human interaction, performance changes might be less likely to be identified. Therefor, performance monitoring plans to control potential causes of undesirable changes in performance should be established to proactively control risk. This plan should ensure that performance changes or degradation is identified and responded to in a timely manner. Deployment of updates, mitigations,

and corrective actions should be also addressed in the plan to ensure a prompt response. Practical examples of risks related AVI data input is addressed by RQ2 in section **4.2.3 CHANGE MANAGEMENT**.

As with any digital or software component, cybersecurity needs to be addressed. Examples of AI/ML specific cybersecurity threats/attacks are included in the draft guidance:

- Data Poisoning: Injecting inauthentic or maliciously modified data
- Model Inversion/stealing: Intentional usage of forged or altered data to infer details from the model or replicate the model that pose risk to intellectual property and privacy breaches
- Model Evasion: Crafting or modifying input samples to deceive models
- Data leakage: Exploitation of vulnerabilities to access sensitive training or inference data
- Overfitting: Deliberate overfit of a model to expose the AI/ML components to adversarial attacks that lead to errors in the model's output
- Model Bias: Manipulation of training data to introduce bias; use of known biases or backdoors in pre-trained models
- Performance drift: Changing or slightly shifting of the underlying data distribution to degrade model performance

For further details on cybersecurity controls and security risk management, the 2023 guidance document "Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions" is referenced (161). This guideline covers topics related to cybersecurity risk management, threat modeling, vulnerability testing, penetration testing, controls to prevent data leakage (access controls, encryption,…) among others.

Lastly, "Appendix E: Example Model Card" of the draft guidance provides an example of a format to provide information of AI/ML models to present data in an order and format that is useful and easy to understand for non-technical audiences. The following AI/ML-specific details are mentioned:

- Model architecture
- Description:
    - Intended users
    - Instructions for use
    - Inputs and outputs of the model
    - Degree of automation
- Performance and Limitations:
    - Accuracy indicators including their 95% two-sided confidence intervals
    - Known biases
    - Precision associated with the outputs
    - Known gaps in the data characterization in development or testing datasets
    - Limitations of the model development or performance evaluation
    - Data characterization such as data sources, data types, sample size, data quality, reference standard and so on
    - Methods used to establish and ensure that the model meets to intended use and user requirements (user acceptance testing, identification of pre-trained models,…)
- Risk Management:
    - Risks associated with the model, data, and outputs (data privacy risks, cybersecurity risks, bias risks, information gaps,…)
    - Interactions, Deployment and Updates (computational resources needed, ongoing performance monitoring, reporting of successes and failures, change management strategies and proactive approaches to address vulnerabilities)
- Development:

- Data used for development of the model.

Appendix F of the draft guidance gives an example of a completed model card for a stand-alone software that contains a ML model that uses a CNN to interpret and analyze ECGs and provides an output on the likelihood of whether the patient has the Disease X and further clinical evaluation is needed.

### 4.4.4 German Notified Bodies Alliance (IG-NB)

The German Notified Bodies Alliance published a questionnaire with best practices from the authors point of view that has undergone several updates over the years, see **TABLE 28**. The questionnaire is partly based on the "Guideline for AI-based medical devices and IVD" Github repository by the Johner-Institute (162).

| Document title | Document type | Reference | Publishing date |
|---|---|---|---|
| IG-NB Fragenkatalog „Künstliche Intelligenz bei Medizinprodukten" | Questionnaire (V2) | (163) | 06.11.2020 |
| | Questionnaire (V3) | (164) | 03.12.2021 |
| IG-NB Questionnaire „Artificial Intelligence (AI) in medical devices" | Questionnaire (V4) | (165) | 09.06.2022 |
| | Questionnaire (V5) | (166) | 15.12.2023 |
| | Questionnaire (V5.1) | (99) | 04.03.2024 |
| IG-NB and Team NB Questionnaire: Artificial Intelligence in Medical Devices | Questionnaire (V1) | (167) | 14.11.2024 |
| | Questionnaire (V1.1) | (168) | 09.12.2024 |

*Table 28 Development of the IG-NB questionnaires on AI in medical devices over the last years.*

The latest version of the questionnaire acknowledges the following limitation in MD certification:

*"Practice has shown that it is difficult for manufacturers to sufficiently prove the conformity for AI devices, which update the underlying models using in-field self-learning mechanisms. Notified bodies do not consider medical devices based on these models to be "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."*

Compared to the US FDAs concept of PCCPs, the European approach is stringent and de-facto excluded continuously-learning medical devices from market access due to the fact that re-certification would be necessary once the performance changes due to self-learning events.

The IG-NB questionnaire formulates its questions around the following structure:

a) General Requirements
b) Requirements for product development
    1. Intended use and stakeholder requirements
    2. Software requirements
    3. Data management
    4. Model development
    5. Product development
    6. Product release
c) Requirements for the post development phases
    1. Production, distribution, installation
    2. Post-Market Surveillance
    3. Decommissioning

The questionnaire offers a detailed, "checklist-style" approach with references to relevant articles of the EU Medical Device Regulation (2017/745), EU In-Vitro Diagnostic Devices Regulation (2017/746), ISO 13485 on Medical Devices and ISO 14971 on Application of risk management to medical devices.

Best practices on competence mapping, for example if the manufacturer outsources software development, can be derived from the questionnaire. For example:

*"3. Competence development*

1. *Does the manufacturer create a list of all roles that are directly or indirectly concerned with AI?*
2. *Does the manufacturer identify AI-related skills for each role (e.g. developers, statisticians, modellers, etc.)?*
3. *Does the manufacturer have adequate records of education, training and competences to conclude that the persons actually have these competences?*
4. *Does the (software) development plans lay out the product-specific competences (beyond or deviating)?*
5. *Is the integration of external competences done according to the rules on outsourced processes? How are outsourced competences recorded/documented?"* (168)

Another section covers valuable questions around labelling procedures of data that is used for model training.

### 4.4.5 Summary (RQ4)

The review of best practices reveals that the medical device industry has rapidly adapted its regulatory framework to address the growing integration of AI/ML technologies under strict standards such as FDA medical device guidelines, EU MDR, IVDR, and ISO 13485. Between 2010 and 2023, hundreds of AI/ML-enabled devices (particularly in radiology) were approved in the United States, underscoring the parallels with the use case of VI of parenteral drug products. Initially, regulatory bodies like the FDA's CDRH provided standalone guidance; however, recent efforts show a shift toward cross-center collaboration and joint guidance with agencies from other countries such as Health Canada and MHRA. Key best-practices include ensuring data integrity, achieving methodological transparency, managing uncertainty in model outputs, and controlling changes due to data drift. To address these topics, frameworks such as ASME V&V 40 have been adapted to assess model credibility based on risk factors like model influence and decision consequence, while concepts like the Predetermined Change Control Plans (PCCPs) facilitate ongoing oversight and iterative improvements without triggering the need for regulatory submissions. Additionally, initiatives like the IG-NB questionnaire in Germany provide detailed checklists to support conformity and competency assessment. Collectively, these best practices offer valuable insights for safe and effective integration of AI/ML in a regulated environment.

## 5 Conclusion

The literature reviewed in this work suggests that the application of AI/ML to VI of parenteral drug products offers significant potential to enhance defect detection and process efficiency, while simultaneously presenting new challenges that must be addressed within GMP environments. The following conclusions can be drawn from research questions one to four (RQ1–RQ4):

AI/ML Use cases in visual inspection (RQ1): The evidence indicates that AI/ML can reduce inspection errors and accelerate the development of vision recipes through techniques such as transfer learning and deep neural network implementation. These approaches have been shown to provide enhanced classification accuracy and improved data-driven insights into the manufacturing process, even though with challenges related to the dataset requirements for model training.

Challenges and opportunities in the GMP setting (RQ2): The integration of AI/ML within a GMP framework needs a robust risk management strategy, comprehensive data governance, and stringent change control measures. Key elements include the adaptation of established frameworks such as ICH

Q9 and GAMP 5 to encompass AI-specific risks, the implementation of computerized system validation (CSV) protocols tailored to AI/ML sub-systems, and the ongoing monitoring of data and cybersecurity. These measures are essential to ensure that data integrity, product quality and patient safety remain uncompromised.

Regulatory Framework and Industry Guidance (RQ3): The evolving legal landscape is characterized by distinct approaches between the European Union and the United States. The EU's focus on hard law to regulate AI is in contrast with the more flexible guidance adopted by the United States of America. Nonetheless, both EMA and FDA underscore the importance of data integrity, model explainability, and continuous monitoring. The insights gained from regulatory initiatives and industry feedback reinforce the necessity for regulatory clarity on the expectations of agencies for the implementation of AI/ML in GMP settings.

Best Practices from the medical device industry (RQ4): A comparative analysis with the medical device industry suggests that best practices - such as iterative change management, the use of transparency tools like data and model cards, and systematic validation frameworks (ASME V&V 40 and Good Machine Learning Practices) – may be directly transferable to pharmaceutical visual inspection. These practices not only facilitate enhanced process control but also support greater regulatory compliance through improved transparency and risk mitigation.

In conclusion, while AI/ML applications in visual inspection of parenteral drug products hold considerable promise for reducing inspection errors, cost and enhancing process efficiency, their successful integration within GMP settings requires a comprehensive, risk-based approach that addresses technical, regulatory, and operational challenges. Future work should focus on refining validation protocols and further adapting cross-industry best practices to ensure that these innovative systems can be implemented safely and effectively in pharmaceutical manufacturing.

## 6 References


1. Shabushnig J. Visual Inspection of Injectable Products: More than Sorting Good from Bad... 2022;

2. Liu F, Hutchinson R. Visible particles in parenteral drug products: A review of current safety assessment practice. Current Research in Toxicology. 2024;7:100175.

3. Bukofzer S, Ayres J, Chavez A, Devera M, Miller J, Ross D, et al. Industry Perspective on the Medical Risk of Visible Particles in Injectable Drug Products. PDA Journal of Pharmaceutical Science and Technology. 2015 Jan 1;69(1):123–39.

4. Shabushnig J. The Visual Inspection Regulatory and Compendial Environment: Current Issues and Opportunities [Internet]. 2024. Available from: https://www.casss.org/docs/default-source/wcbp/2024-speaker-presentations/shabushnig-john-g.-insight-pharma-consulting-llc-2024.pdf?sfvrsn=fe46030a_6

5. Shabushnig J, Miller R, Watson R. PDA Research: 2023 PDA Visual Inspection Survey.

6. Körber Pharma [Internet]. [cited 2025 Feb 21]. AI-basierte Inspektion. Available from: https://www.koerber-pharma.com/solutions/inspektion/automatische-inspektion/ai-basierte-inspektion

7. Syntegon introduces fully integrated syringe inspection line with AI [Internet]. 2021 [cited 2025 Feb 21]. Available from: https://www.syntegon.com/press/fully-integrated-syringe-inspection-line

8. Brevetti AI - Visual Inspection [Internet]. [cited 2025 Feb 21]. Available from: https://brevetti.ai/information?view=article&id=6&catid=8

9. Artificial Intelligence - Stevanato Group [Internet]. [cited 2025 Feb 21]. Available from: https://www.stevanatogroup.com/en/technologies-equipment/visual-inspection/inspection-technologies/artificial-intelligence/

10. Straub J. Practical Applications, Challenges and Solutions When Using Deep Learning and Big Data for Automated Visual Inspection [Internet]. PDA Visual Inspection Forum; 2022 Apr; Berlin. Available from: https://www.pda.org/docs/default-source/attendee-presentations/europe/2022/04_vi_forum/straub_joseph.pdf

11. Gaulding S, Stauffer R. PDA's 4th Metrics Conference: Measuring the State of Quality Metrics.

12. Committee to Identify Innovative Technologies to Advance Pharmaceutical Manufacturing, Board on Chemical Sciences and Technology, Division on Earth and Life Studies, National Academies of Sciences, Engineering, and Medicine. Innovations in Pharmaceutical Manufacturing on the Horizon: Technical Challenges, Regulatory Issues, and Recommendations [Internet]. Washington, D.C.: National Academies Press; 2021 [cited 2024 Feb 3]. Available from: https://www.nap.edu/catalog/26009

13. European Commission. EU Good Manufacturing Practice (GMP) Annex 1: Manufacture of Sterile Medicinal Products [Internet]. 2022. Available from: https://health.ec.europa.eu/document/download/e05af55b-38e9-42bf-8495-194bbf0b9262_en?filename=20220825_gmp-an1_en_0.pdf

14. European Commission. EU Good Manufacturing Practice (GMP) Annex 11: Computerised Systems [Internet]. 2011 [cited 2024 Oct 27]. Available from: https://www.taylorfrancis.com/books/9780203026656/chapters/10.3109/9780203026656-14

15. AI Act - REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL aying 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). [Internet]. 2024 [cited 2025 Feb 28]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

16. Knapp JZ, Kushner HK. Generalized Methodology For Evaluation of Parenteral Inspection Procedures. PDA J Pharm Sci Technol. 1980 Jan 1;34(1):14.

17. Delgado J. Deep Learning Development and Qualification in Automated Vision Inspection Technology for Parenteral Pharmaceutical Drug Products [Internet]. PDA Visual Inspection Forum; 2019 Apr; Washington, D.C. Available from: https://www.pda.org/docs/default-source/attendee-presentations/north-america/2019/2019-pda-visual-inspection-forum/jorge-delgado.pdf

18. Veillon R, Shabushnig J, Aabye-Hansen L, Duvinage M, Eckstein C, Li Z, et al. Applying Machine Learning to the Visual Inspection of Filled Injectable Drug Products. PDA Journal of Pharmaceutical Science and Technology. 2023 Jun 15;pdajpst.2022.012796.

19. Veillon R. Human Like Vision Robot. PDA Visual Inspection Forum; 2022 Apr; Berlin.

20. Veillon R. Transforming Pharmaceutical Inspections with Human-Like Robot AI [Internet]. 2024 [cited 2025 Feb 21]. Available from: https://www.pda.org/pda-letter-portal/home/full-article/transforming-pharmaceutical-inspections-with-human-like-robot-ai

21. Liang Q, Luo B. Visual inspection intelligent robot technology for large infusion industry. Open Computer Science. 2023 Feb 8;13(1):20220262.

22. Sardella A. Latest technology - Deep learning solutions for visual inspection of parenteral. Human-like inspection for personalized therapies. The Universe of Pre-filled Syringes and Injection Devices; 2019 23.10; Gothenburg, Sweden.

23. Adelphi [Internet]. [cited 2025 Feb 26]. Apollo II Liquid Viewer I Adelphi Manufacturing. Available from: https://www.adelphi.uk.com/at/product/apollo-ii-liquid-inspection-unit/

24. Körber Pharma [Internet]. [cited 2025 Feb 26]. Halbautomatische Inspektion V90+. Available from: https://www.koerber-pharma.com/solutions/inspektion/halbautomatische-inspektion-v90

25. Brevetti K15 DR - Brevetti CEA [Internet]. [cited 2025 Feb 26]. Available from: https://www.brevetti-cea.com/machines/brevetti-k15-dr

26. Vision Robot Unit (VRU) - Stevanato Group [Internet]. [cited 2025 Feb 26]. Available from: https://www.stevanatogroup.com/en/technologies-equipment/visual-inspection/vision-robot-unit/

27. Heynes C, Lumkemann J. X-Ray Detection of Foreign Particles in Freeze Dried Products. Freeze drying conference; 2014; Brussels, Belgium.

28. Sacher S, Poms J, Dekner M, Wallner-Mang S, Vogt M, Khinast JG, et al. X-ray imaging: A potential enabler of automated particulate detection and cake-structure analysis in lyophilized products? International Journal of Pharmaceutics: X. 2021 Dec;3:100101.

29. Cherris R, Valley U, Aabye-Hansen L, McLean R, Overroedder D, Shabushnig J, et al. PDA Technical Report No. 79 - Particulate Matter Control in Difficult to Inspect Parenterals [Internet]. 2018. Available from: https://www.pda.org/bookstore/product-detail/4246-tr-79-particulate-matter-inspect-parenterals

30. USP〈1790〉 Visual Inspection of Injections [Internet]. [cited 2024 Feb 3]. Available from: https://doi.usp.org/USPNF/USPNF_M7198_06_01.html

31. Melchore JA. Sound Practices for Consistent Human Visual Inspection. AAPS PharmSciTech. 2011 Mar;12(1):215–21.

32. McGee A. Visual Inspection - Regulatory Trends. Challenges, and Ovservations during Regulatory Inspections. PDA Ireland Event; 2019 Nov 22.

33. Visual acuity. In: Wikipedia [Internet]. 2025 [cited 2025 Feb 22]. Available from: https://en.wikipedia.org/w/index.php?title=Visual_acuity&oldid=1271708998

34. McDonagh A. Manual Inspection Capabilities and Best Practices, including Knapp Study. 2019 PDA Ireland Visual Inspection Event; 2019 Nov 22.

35. Cherris RT. Visual Inspection Lifecycle - Particulate and Container/Closure Defects [Internet]. 2013 Aug. Available from: http://www.baipharma.com/documents/1-Cherris-InspectionLifecycle-PDAVIForumOct2011-ExpandedAug2013.pdf

36. McLean RW. Case Study Comparison of Qualification Data for Automated Visual Inspection.

37. Berdovich D. Understanding the Variables When Creating and Using Challenge Sets to Qualify Inspection Methods. PDA Forum on Visual Inspection; 2013 Oct 22.

38. Posset T, Gaus H, Dearden M, Becker M, Goodwin A, Eicher R. ECA Visual Inspection of medicinal products for parenteral use - Good Practice Guide [Internet]. 2020. Available from: https://visual-inspection.gmp-compliance.org/

39. Burazer A. Understanding Japan Quality [Internet]. 2024 [cited 2025 Feb 21]. Available from: https://www.pda.org/pda-letter-portal/home/full-article/understanding-japan-quality

40. FDA Draft Guidance: Inspection of Injectable Products for Visible Particulates - Guidance for Industry [Internet]. 2021. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/inspection-injectable-products-visible-particulates

41. Wang SH. Regulatory Perspective on Inspection of Injectable Products for Visible Particulates [Internet]. 2020. Available from: https://pda-asiapacific.glueup.com/resources/protected/organization/1176/event/26285/90366038-46e0-4b68-847a-fd73a231dd98.pdf

42. Langille SE. Visible Particulate Contamination Control for Injectable Products: A Life-Cycle Approach. PDA J Pharm Sci Technol. 2020;74(3):359–66.

43. Essentially Free: How A Definition Tamed The Visible Particulate Matter Debate [Internet]. [cited 2025 Feb 21]. Available from: https://www.outsourcedpharma.com/doc/essentially-free-how-a-definition-tamed-the-visible-particulate-matter-debate-0001

44. Glassner A. Deep Learning: A Visual Approach. No Starch Press; 2021.

45. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006. 738 p. (Information science and statistics).

46. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Mass: The MIT press; 2016. (Adaptive computation and machine learning).

47. Deep learning. In: Wikipedia [Internet]. 2025 [cited 2025 Feb 23]. Available from: https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=1275051593

48. Neural network (machine learning). In: Wikipedia [Internet]. 2025 [cited 2025 Feb 23]. Available from: https://en.wikipedia.org/w/index.php?title=Neural_network_(machine_learning)&oldid=1276196615

49. Convolutional neural network. In: Wikipedia [Internet]. 2025 [cited 2025 Feb 23]. Available from: https://en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=1276806099

50. Recurrent neural network. In: Wikipedia [Internet]. 2025 [cited 2025 Feb 23]. Available from: https://en.wikipedia.org/w/index.php?title=Recurrent_neural_network&oldid=1274774444

51. FDA. FDA Digital Health and Artificial Intelligence Glossary – Educational Resource. FDA [Internet]. 2024 Sep 26 [cited 2024 Oct 19]; Available from: https://www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource

52. European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle (final) [Internet]. 2024. Available from: https://www.ema.europa.eu/en/use-artificial-intelligence-ai-medicinal-product-lifecycle

53. Kazmi M, Hafeez B, Aftab F, Shahid J, Qazi SA. A Deep Learning-Based Framework for Visual Inspection of Plastic Bottles. IEEE Access. 2023;11:125529–42.

54. Zoph B, Le QV. Neural Architecture Search with Reinforcement Learning [Internet]. arXiv; 2017 [cited 2025 Feb 23]. Available from: http://arxiv.org/abs/1611.01578

55. Jia Z, Wang M, Zhao S. A review of deep learning-based approaches for defect detection in smart manufacturing. J Opt. 2024 Apr;53(2):1345–51.

56. Defect Detection Model Trained on Synthetic Data w/ Omniverse Code, Adobe Substance 3D, and Roboflow [Internet]. 2023 [cited 2025 Feb 23]. Available from: https://www.youtube.com/watch?v=cuoZAYoewE0

57. Park M, Jeong J. Design and Implementation of Machine Vision-Based Quality Inspection System in Mask Manufacturing Process. Sustainability. 2022 May 16;14(10):6009.

58. Alstrup I, Danish Medicines Agency (DKMA). Questions to critical GxP AI-ML applications v. 0.9.4.

59. Type I and type II errors. In: Wikipedia [Internet]. 2025 [cited 2025 Mar 30]. Available from: https://en.wikipedia.org/w/index.php?title=Type_I_and_type_II_errors&oldid=1282365241

60. Bias–variance tradeoff. In: Wikipedia [Internet]. 2025 [cited 2025 Mar 2]. Available from: https://en.wikipedia.org/w/index.php?title=Bias%E2%80%93variance_tradeoff&oldid=1276537347

61. Oztemel E. Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management; Chapter 1: Intelligent Manufacturing Systems. 2010. (Springer Series in Advanced Manufacturing).

62. Veillon R, Koert F, Koch S. Mastering AVI - Module 9: Future trends in AVI [Internet]. 2022. Available from: https://www.pda.org/docs/default-source/pda-europe-education-presentations/2022/04_pda-eu00134-mastering-automated-to-vi/module9_future_trends_in-_avi_mastering_avi.pdf

63. Riehn F. Challenges from a machine supplier's perspective. PDA EU00186 How to Handle Test Sets in Visual Inspection; 2024 Feb 13; Markt Schwaben, Germany.

64. Tsay C, Li Z. Automating Visual Inspection of Lyophilized Drug Products With Multi-Input Deep Neural Networks. In: 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE) [Internet]. Vancouver, BC, Canada: IEEE; 2019 [cited 2024 Feb 3]. p. 1802–7. Available from: https://ieeexplore.ieee.org/document/8843069/

65. Abdul-Fattah AM, Oeschger R, Roehl H, Bauer Dauphin I, Worgull M, Kallmeyer G, et al. Investigating factors leading to fogging of glass vials in lyophilized drug products. European Journal of Pharmaceutics and Biopharmaceutics. 2013 Oct;85(2):314–26.

66. Zhao M, Zhang H, Liu L, Liang Z, Deng G. Joint Deep Learning and Clustering Algorithm for Liquid Particle Detection of Pharmaceutical Injection. In: 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA) [Internet]. Xi'an: IEEE; 2018 [cited 2024 Feb 3]. p. 1–6. Available from: https://ieeexplore.ieee.org/document/8608158/

67. Goodwin A, Brandenburger. Deep Learning for Enhanced AVI Detection combined with Rule-Based Image Sequence Tracking [Internet]. PDA Visual Inspection Forum; 2024 Apr; Munich. Available from: https://www.pda.org/docs/default-source/attendee-presentations/europe/2024/04_visual/goodwin-al-and-brandenburger-wolfgang.pdf

68. Tran P. Automated Visual Inspection of Lyophilized Products via Deep Learning and Autoencoders. Massachusetts Institute of Technology; 2021.

69. Hollander M. Evaluating Future Biopharmaceutical Inspection Needs, Infrastructure Capability Gaps, and Technology Development Strategies [Internet]. 2012. Available from: https://dspace.mit.edu/handle/1721.1/74919

70. Martinez E. The theory behind automatic inspection technologies for subvisible-to-visible particle detection and container closure integrity. 2016 Nov 9.

71. He X, Zhao K, Chu X. AutoML: A survey of the state-of-the-art. Knowledge-Based Systems. 2021 Jan;212:106622.

72. Google Cloud [Internet]. [cited 2025 Feb 28]. AutoML Solutions - Train models without ML expertise. Available from: https://cloud.google.com/automl

73. Amazon Web Services, Inc. [Internet]. [cited 2025 Feb 28]. Machine Learning ohne Code – Amazon SageMaker Canvas – AWS. Available from: https://aws.amazon.com/de/sagemaker-ai/canvas/

74. Azure AI Custom Vision | Microsoft Azure [Internet]. [cited 2025 Feb 28]. Available from: https://azure.microsoft.com/en-us/products/ai-services/ai-custom-vision

75. Clive Humby. In: Wikipedia [Internet]. 2024 [cited 2025 Feb 28]. Available from: https://en.wikipedia.org/w/index.php?title=Clive_Humby&oldid=1253087022

76. Veillon R. Mastering AVI - Module 2: Introduction to technical principles of automated inspection machines [Internet]. PDA; 2022. Available from: https://www.pda.org/docs/default-source/pda-europe-education-presentations/2022/04_pda-eu00134-mastering-automated-to-vi/module2_avi_equipment_pda_mastering_automated_visual_inspection.pdf

77. Zahel T. Digitale Zwillinge von Gesamtprozessen. Pharm Ind. 2024;86(7):695–9.

78. Yadav V, Kennedy C. Vision Inspection using Machine Learning/Artificial Intelligence. Pharmaceutical Engineering. 2020 Dec;40(6):34–9.

79. Application of Artificial Intelligence AI & Machine Learning ML in pharmaceutical Industry [Internet]. Youtube; 2022 [cited 2024 Sep 1]. (Pharma Best Practices Webinars). Available from: https://youtu.be/NhyRPVszj0A?si=YPJsie1PApBZ2-XT

80. Straub J. Collection, Archival, and Analysis of Images from Automated Vision Inspection: Present Value and Prospective Opportunities [Internet]. PDA Visual Inspection Forum; 2019 Apr; Washington, D.C. Available from: https://www.pda.org/docs/default-source/attendee-presentations/north-america/2019/2019-pda-visual-inspection-forum/joseph-a-straub.pdf

81. Tellbach D. Reducing Inspection Complexity by Introducing Beyond-Human Sensing for Lyophilized Vaccine Inspection [Internet]. PDA Visual Inspection Forum; 2024 Apr; Munich. Available from: https://www.pda.org/docs/default-source/attendee-presentations/europe/2024/04_visual/tellbach-denise.pdf

82. Holl M, Corbisiero A. Unlocking Potential: AI-Powered Visual Inspection in Commercial Use [Internet]. PDA Visual INspection Forum; 2024 Apr; Munich. Available from: https://www.pda.org/docs/default-source/attendee-presentations/europe/2024/04_visual/holl-mario-and-corbisiero-alessandra.pdf

83. Tabassi E. Artificial Intelligence Risk Management Framework (AI RMF 1.0) [Internet]. Gaithersburg, MD: National Institute of Standards and Technology (U.S.); 2023 Jan [cited 2024 Feb 3] p. NIST AI 100-1. Report No.: NIST AI 100-1. Available from: http://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

84. ISPE GAMP RDI Good Practice Guide: Data Integrity by Design [Internet]. ISPE; 2020. Available from: https://ispe.org/publications/guidance-documents/gamp-rdi-good-practice-guide-data-integrity-design

85. ISPE GAMP 5: a risk-based approach to compliant gxp computerized systems [Internet]. ISPE; 2022. Available from: https://ispe.org/publications/guidance-documents/gamp-5-guide-2nd-edition

86. Blumenthal R, Erdmann N, Heitmann M, Lemettinen AL, Stockton B. Machine Learning Risk and Control Framework. Pharmaceutical Engineering. 2024 Feb;44(1):12–22.

87. Erdmann N, Blumenthal R, Baumann I, Kaufmann M. AI Maturity Model for GxP Application: A Foundation for AI Validation. Pharmaceutical Engineering. 2022 Apr;18–24.

88. Staib EJ, Williams TG, Wyn S. Applying GAMP® Concepts to Machine Learning. Pharmaceutical Engineering. 2023 Feb;43(1):14–23.

89. Altrabsheh E, Heitmann M, Lochbronner A. AI Governance and QA Framework: AI Governance Process Design. Pharmaceutical Engineering. 2022 Aug;42(4):47–53.

90. Brindle A, Davy S, Tiffany D, Watts C. Risk Analysis and Mitigation Matrix (RAMM) - A Risk Tool for Quality Management. Pharmaceutical Engineering. 2012 Feb;32(1):26–33.

91. ICH Q9(R1) Quality Risk Management [Internet]. Available from: https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_Step4_2023_0126_0.pdf

92. Veillon R. Automated Visual Inspection Deep Learning Case Study [Internet]. FDA/PQRI Workshop; 2023 Sep 26 [cited 2024 Sep 10]. Available from: https://pqri.org/wp-content/uploads/2023/09/6-Veillon-Automated-Visual-Inspection-Deep-Learning-Case-Study-final-01-sept-2023.pdf

93. Panza R. Artificial Intelligence Validation approach. GAMP UK Forum; 2020 Apr 21.

94. FDA. Artificial Intelligence in Drug Manufacturing [Internet]. 2023. Available from: https://www.fda.gov/media/165743/download

95. ISPE GAMP® Good Practice Guide: IT Infrastructure Control and Compliance (Second Edition). 2017 Aug; Available from: https://ispe.org/publications/guidance-documents/gamp-it-infrastructure-control-compliance

96. Sahiner B, Friedman B, Linville C, Ipach C, Montgomery E. Perspectives and Best Practices for Artificial Intelligence and Continuously Learning Systems in Healthcare. In Cincinnati; 2018.

97. Machine Learning Best Practices in Healthcare and Life Sciences - AWS Whitepaper.

98. GxP Systems on AWS [Internet]. 2021. Available from: https://d1.awsstatic.com/whitepapers/compliance/Using_AWS_in_GxP_Systems.pdf

99. IG-NB. Questionnaire Artificial Intelligence in medical devices - Version 5.1 [Internet]. 2024. Available from: https://www.ig-nb.de/index.php?eID=dumpFile&t=f&f=3692&token=34f73f670e9d31f5890efcfe4434d2a935a57ab1

100. Su J, Vargas DV, Kouichi S. One pixel attack for fooling deep neural networks. IEEE Trans Evol Computat. 2019 Oct;23(5):828–41.

101. Kriesel D. Glaube keinem Scan, den du nicht selbst gefälscht hast [Internet]. Chaos Communcation Congress - 31C3; 2014 Dec 28; Hamburg. Available from: https://www.dkriesel.com/_media/blog/2014/xerox-ccc-davidkriesel.pdf

102. Herm LV, Heinrich K, Wanner J, Janiesch C. Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. International Journal of Information Management. 2023 Apr;69:102538.

103. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. 2020 Feb;128(2):336–59.

104. Lakshmanan V, Görner M, Gillard R. Practical Machine Learning for Computer Vision. 1st ed. O'Reilly Media; 2021.

105. Altrabsheh E, Heitmann M, Steinmüller P, Pastory Vinco B. The Road to Explainable AI in GxP-Regulated Areas. Pharmaceutical Engineering. 2023 Feb;43(1):24–32.

106. Hybrid Public Workshop on Artificial Intelligence in Drug & Biological Product Development [Internet]. 2024 [cited 2025 Feb 3]. Available from: https://vimeo.com/1001681072

107. REdI Annual Conference 2024: CDER (Drugs) Innovation in Medical Product Development (Day 1 of 2) [Internet]. [cited 2024 Oct 6]. Available from: https://www.youtube.com/watch?v=Yj9i2QRSVM4

108. The possible Impact of Artificial Intelligence on the Role of the Qualified Person - ECA Academy [Internet]. [cited 2025 Mar 1]. Available from: https://www.gmp-compliance.org/gmp-news/the-possible-impact-of-artificial-intelligence-on-the-role-of-the-qualified-person

109. Mintanciyan A, Budihandojo R, English J, Lopez O, Matos JE, McDowall R. Artificial Intelligence Governance in GxP Environments. Pharmaceutical Engineering. 2024 Aug;44(4):62–7.

110. Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten (ZLG). Aide-Mémoire 07121203 Überwachung computergestützter Systeme [Internet]. 2022 [cited 2024 Oct 19]. Available from: https://www.zlg.de/index.php?eID=dumpFile&t=f&f=2601&token=587fd3625e2d3b9ad2956f334ef67bdd913123eb

111. Pharmaceutical Inspection Co-Operation Scheme (PIC/S). Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments [Internet]. 2021. Available from: https://picscheme.org/docview/4234

112. Food and Drug Administration (FDA). Data Integrity and Compliance with Drug CGMP: Questions and Answers. Guidance for Industry. Questions and Answers [Internet]. 2018 Dec; Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-integrity-and-compliance-drug-cgmp-questions-and-answers

113. FDA. Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations: Questions and Answers [Internet]. 2024. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/electronic-systems-electronic-records-and-electronic-signatures-clinical-investigations-questions

114. European Commission. EudraLex - Volume 4 - Good Manufacturing Practice (GMP) guidelines. 2011 [cited 2025 Jul 13]. EU Good Manufacturing Practice (GMP) Volume 4, Part I - Basic Requirements for Medicinal Products: Chapter 4 - Documentation. Available from: https://www.taylorfrancis.com/books/9780203026656/chapters/10.3109/9780203026656-14

115. EMA. Guideline on computerised systems and electronic data in clinical trials [Internet]. 2023. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf

116. European Medicines Agency (EMA). Guidance on good manufacturing practice and good distribution practice: Questions and answers | European Medicines Agency (EMA) [Internet].

2021 [cited 2024 Oct 6]. Available from: https://www.ema.europa.eu/en/human-regulatory-overview/research-development/compliance-research-development/good-manufacturing-practice/guidance-good-manufacturing-practice-good-distribution-practice-questions-answers

117. Medicines & Healthcare products Regulatory Agency (MHRA). GXP Data Integrity Guidance and Definitions [Internet]. 2018. Available from: https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity

118. World Health Organization (WHO). TRS 1033 - Annex 4: WHO Guideline on data integrity [Internet]. 2021. Available from: https://www.who.int/publications/m/item/annex-4-trs-1033

119. Ullagaddi P. Safeguarding Data Integrity in Pharmaceutical Manufacturing. J Adv Med Pharm Sci. 2024 Aug 17;26(8):64–75.

120. Holland S, Hosny A, Newman S, Joseph J, Chmielinski K. The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. 2020 [cited 2024 Oct 5]; Available from: http://www.bloomsburycollections.com/book/data-protection-and-privacy-data-protection-and-democracy

121. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé III H, et al. Datasheets for Datasets [Internet]. arXiv; 2021 [cited 2024 Feb 3]. Available from: http://arxiv.org/abs/1803.09010

122. Pushkarna M, Zaldivar A, Kjartansson O. Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In: 2022 ACM Conference on Fairness, Accountability, and Transparency [Internet]. Seoul Republic of Korea: ACM; 2022 [cited 2024 Sep 29]. p. 1776–826. Available from: https://dl.acm.org/doi/10.1145/3531146.3533231

123. High-Level Expert Group on Artificial Intelligence (AI HLEG) by the European Commission. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. 2020 Jul 17; Available from: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

124. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning [Internet]. arXiv; 2022 [cited 2024 Oct 6]. Available from: http://arxiv.org/abs/1908.09635

125. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds & Machines. 2018 Dec;28(4):689–707.

126. Chmielinski K, Newman S, Kranzinger C, Hind M, Wortman Vaughan J, Mitchell M, et al. The CLeAR Documentation Framework for AI Transparency: Recommendations for Practitioners and Context for Policymakers [Internet]. 2024 [cited 2024 Oct 6]. Available from: https://shorensteincenter.org/wp-content/uploads/2024/05/CleAR_KChmielinski_FINAL.pdf

127. Burkart N, Huber MF. A Survey on the Explainability of Supervised Machine Learning. jair. 2021 Jan 19;70:245–317.

128. Panza R. Artificial Intelligence Validation approach. GAMP UK Forum; 2019 Oct 9; Macclesfield.

129. Stockton B. Use of a Control Framework to Limit Risk of AI in GxP [Internet]. Boston GAMP CoP Forum; 2024 Jun 6; Boston. Available from:

https://www.ispeboston.org/download/educational_presentations/2024/Use-of-a-Control-Framework-to-Limit-risk-of-AI-in-GXP_20240606-StocktonB-Final-20240601.pdf

130. Ramasesh V, Lewkowycz A, Dyer E. EFFECT OF MODEL AND PRETRAINING SCALE ON CATASTROPHIC FORGETTING IN NEURAL NETWORKS. 2022;

131. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model Cards for Model Reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency [Internet]. 2019 [cited 2024 Aug 17]. p. 220–9. Available from: http://arxiv.org/abs/1810.03993

132. Artificial Intelligence Act. In: Wikipedia [Internet]. 2024 [cited 2024 Oct 27]. Available from: https://en.wikipedia.org/w/index.php?title=Artificial_Intelligence_Act&oldid=1249744296

133. EMA GMP/GDP IWG, PIC/S. Concept paper on the revision of GMP Annex 11 – Computerised Systems [Internet]. 2022. Available from: https://picscheme.org/docview/4967

134. European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle (draft) [Internet]. 2023. Available from: https://www.ema.europa.eu/en/use-artificial-intelligence-ai-medicinal-product-lifecycle#document-history-70150

135. HMA-EMA Big Data Steering Group. Multi-annual AI workplan 2023-2028 Version 1 [Internet]. 2023 [cited 2024 Oct 27]. Available from: https://www.ema.europa.eu/en/documents/work-programme/multi-annual-artificial-intelligence-workplan-2023-2028-hma-ema-joint-big-data-steering-group_en.pdf

136. European Commission. Stakeholders' Consultation on EudraLex Volume 4 - Good Manufacturing Practice Guidelines: Chapter 4, Annex 11 and New Annex 22. 2025 [cited 2025 Jul 13]. EU Good Manufacturing Practice (GMP) Volume 4 Draft guideline: New Annex 22 - Artificial Intelligence. Available from: https://health.ec.europa.eu/document/download/5f38a92d-bb8e-4264-8898-ea076e926db6_en?filename=mp_vol4_chap4_annex22_consultation_guideline_en.pdf

137. FDA. FDA Discussion Paper and Request for Feedback: Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products. 2023 May 10;

138. FDA Artificial Intelligence & Medicinal Products: How CBER, CDER, CDRH, and OCP are Working Togehter. 2024 Mar [cited 2025 Feb 3]; Available from: https://content.govdelivery.com/attachments/USFDA/2024/03/15/file_attachments/2815628/omp_aimedicalproducts_final_240313%20%281%29.pdf

139. FDA. FDA Draft Guidance: Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products. FDA. 2025 Jan 6;

140. Joint Heads of Medicines Agencies (HMA)/ (EMA) AI workshop – Smart regulation – 20 November 2023 [Internet]. 2023 [cited 2025 Feb 4]. Available from: https://www.youtube.com/watch?v=t-mkpIlHSL0

141. The White House [Internet]. [cited 2025 Feb 28]. Blueprint for an AI Bill of Rights | OSTP. Available from: https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

142. Quality Innovation Group | European Medicines Agency (EMA) [Internet]. 2022 [cited 2024 Oct 27]. Available from: https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp-working-parties-other-groups/quality-innovation-group

143. European Medicines Agency (EMA). Implementation of comments received on the draft "Reflection paper on the use of artificial intelligence in the medicinal product lifecycle" (EMA/CHMP/CVMP/83833/2023) [Internet]. 2024 [cited 2024 Oct 27]. Available from: https://www.ema.europa.eu/en/documents/comments/implementation-comments-received-draft-reflection-paper-use-artificial-intelligence-medicinal-product-lifecycle-ema-chmp-cvmp-83833-2023_en.pdf

144. European Commission. Stakeholders' Consultation on EudraLex Volume 4 - Good Manufacturing Practice Guidelines: Chapter 4, Annex 11 and New Annex 22. 2025 [cited 2025 Jul 13]. EU Good Manufacturing Practice (GMP) Volume 4 Draft guideline: Revised Chapter 4 - Documentation. Available from: https://health.ec.europa.eu/document/download/fa336a6e-753b-46fc-b53b-9178bacd8878_en?filename=mp_vol4_chap4_consultation_guideline_en.pdf

145. European Commission. Stakeholders' Consultation on EudraLex Volume 4 - Good Manufacturing Practice Guidelines: Chapter 4, Annex 11 and New Annex 22. 2025 [cited 2025 Jul 13]. EU Good Manufacturing Practice (GMP) Volume 4 Draft guideline: Revision of Annex 11 – Computerised Systems. Available from: https://health.ec.europa.eu/document/download/40231f18-e564-4043-94de-c031f813d38b_en?filename=mp_vol4_chap4_annex11_consultation_guideline_en.pdf

146. FDA. Emerging Technology Program (ETP). FDA [Internet]. 2025 Jul 1 [cited 2025 Feb 3]; Available from: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/emerging-technology-program-etp

147. Federal Register [Internet]. 2025 [cited 2025 Feb 5]. Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry; Availability; Comment Request. Available from: https://www.federalregister.gov/documents/2025/01/07/2024-31542/considerations-for-the-use-of-artificial-intelligence-to-support-regulatory-decision-making-for-drug

148. Fakhouri T. Responsive Regulation of Artificial Intelligence in Drug Development. Regulatory Education for Industry (REdI) Conference 2024. 2024 May 29;

149. BioPhorum, Opincar M, Smiatek J, Presser B, Lemieux L, Johnston A, et al. Biophorum response to Danish agency [Internet]. BioPhorum; 2021 Sep [cited 2024 Feb 12]. Available from: https://www.biophorum.com/download/regulatory-feedback-to-dmka-questions-to-critical-gxp-augmented-intelligence/

150. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. FDA [Internet]. 2024 Dec 20 [cited 2025 Feb 9]; Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

151. FDA Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021.

152. FDA Discussion Paper: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD).

153. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions - Guidance for Industry and Food and Drug Administration Staff. 2024 Dec 4;

154. FDA. Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. 2025 Jan 7; Available from: https://www.regulations.gov/document/FDA-2024-D-4488-0002

155. ASME V&V 40 Webinar [Internet]. 2024 [cited 2025 Mar 1]. Available from: https://www.youtube.com/watch?v=2Yo7zFZfdKA

156. U.S. Food and Drug Administration (FDA), Health Canada, United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA). Good Machine Learning Practice (GMLP) for Medical Device - Guiding Principles [Internet]. 2021. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

157. U.S. Food and Drug Administration (FDA), Health Canada, United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA). Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles

158. U.S. Food and Drug Administration (FDA), Health Canada, United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA). Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles. 2024 Jun; Available from: https://www.fda.gov/media/179269/download?attachment

159. Stueben DJ, Heitmann M, Baumann I. Good Machine Learning Practices in GxP regulated areas. Pharm Ind. 2024;

160. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions (DRAFT). 2023 Apr 3; Available from: https://www.regulations.gov/document/FDA-2022-D-2628-0002

161. FDA. Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions. 2023 Sep 27; Available from: https://www.fda.gov/media/119933/download

162. Johner Institut. GitHub. [cited 2025 Mar 1]. Guideline for AI-based medical devices and IVD. Available from: https://github.com/johner-institut/ai-guideline/blob/master/Guideline-AI-Medical-Devices_EN.md

163. IG-NB. Fragenkatalog Künstliche Intelligenz bei Medizinprodukten - Version 2 [Internet]. 2020. Available from: https://www.ig-nb.de/index.php?eID=dumpFile&t=f&f=1507&token=4be924185b49975aaa712786f0729f831134d4d5

164. IG-NB. Fragenkatalog Künstliche Intelligenz bei Medizinprodukten - Version 3 [Internet]. 2021. Available from: https://www.ig-nb.de/index.php?eID=dumpFile&t=f&f=2417&token=dfc3afb59d7b85af65ffab53dfaf667af094aebb

165. IG-NB. Questionnaire Artificial Intelligence in medical devices - Version 4 [Internet]. 2022. Available from: https://www.ig-nb.de/index.php?eID=dumpFile&t=f&f=2618&token=010db38d577b0bfa3c909d6f1d74b19485e86975

166. IG-NB. Questionnaire Artificial Intelligence in medical devices - Version 5 [Internet]. 2023. Available from: https://www.ig-nb.de/index.php?eID=dumpFile&t=f&f=3617&token=52b6309803b29f4710928b64ce457652f45be0e5

167. German Notified Bodies Alliance for Medical Devices (IG-NB), The European Association of Medical Devices Notified Bodies (Team NB). Questionnaire Artificial Intelligence in medical devices - Version 1. 2024 Nov 14; Available from: https://www.team-nb.org/wp-content/uploads/2024/11/Team-NB-PositionPaper-AI-in-MD-Questionnaire-V1-20241125.pdf

168. German Notified Bodies Alliance for Medical Devices (IG-NB), The European Association of Medical Devices Notified Bodies (Team NB). Questionnaire Artificial Intelligence in medical devices - Version 1.1. Artificial Intelligence. 2024 Dec 9;

## 7 Appendix A: List of Illustrations

## 8 Appendix B: List of Tables

## 9 Abbreviations and Definitions (Glossary)

| Acronym / Term | Definition |
|---|---|
| Adversarial attack | An attack technique that introduces crafted input data to a model with the intent to generate incorrect outputs. |
| AI | Artificial intelligence, refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. |
| AI Performance Monitoring | "Refers to the process of regularly collecting and analyzing data on the use of a deployed AI system to evaluate its performance in accomplishing its intended tasks in real-world settings. The assessment of an AI model's performance involves various performance metrics and criteria depending on the specific application. This monitoring typically aims to assess the performance of these AI systems in practice, detect performance degradation or changes (e.g., due to data drift), identify instances of misuse, and address any safety or usability concerns" (51) |
| AI RAMM | Artificial Intelligence Risk Analysis and Mitigation Matrix |
| AI/ML | Artificial Intelligence / Machine Learning |
| AI-DSF | AI-enabled device software functions(153) |
| AISC | AI Steering Committee. The AISC was established by CDER to coordinate efforts around AI uses across therapeutic development |
| ALCOA | Attributable, Legible, Contemporaneous, Original, and Accurate. ALCOA represents principles to implement data integrity. |
| ALCOA++ | An expansion of the ALCOA principles. + Complete + Consistent + Enduring + Available |
| ANN | Artificial Neural Network |
| AQL | Acceptable Quality Level |
| ASME | American Society of Mechanical Engineers |
| AVI | Automated Visual Inspection |
| AWS | Amazon Web Services |
| AZ | Accept Zone (Knapp methodology) |
| BBox | Bounding Box |
| Bias | A systematic error from oversimplified assumptions in a model that prevents it from accurately capturing the underlying data patterns. |
| Black box | A system which can be viewed in terms of its inputs and outputs without any knowledge of its internal workings. |
| Catastrophic forgetting | Tendency of an ANN to abruptly forget previously learned information upon learning new information. |
| CATT | CBERs Advanced Technologies Team |
| CBER | Center for Biologics Evaluation and Research (US FDA) |
| CCS | Contamination Control Strategy |
| CDER | Center for Drug Evaluation and Research (US FDA) |
| CNN | Concolutional Neural Network |
| Confusion matrix | A table that is used to define the performance of an algorithm. |
| COTS | Commercial off-the-shelf |
| CQA | Critical Quality Attribute |
| CS | Computerized System |
| CSV | Computerized System Validation |
| CV | Computer Vision |
| Data Augmentation | A set of techniques to artificially expand the size and diversity of a training dataset. |
| Data Card | "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" (51) |
| Data Drift | "Refers to the change in the input data distribution a deployed model receives over time, which can cause the model's performance to degrade. This occurs when the properties of the underlying data change. Data drift can affect the accuracy and reliability of predictive models" (51) |

| Acronym / Term | Definition |
|---|---|
| Data Leakage | In Machine Learning, information from one dataset leaks into another dataset. For example: If duplicate data is present in the training as well as the test dataset. One way to avoid data leakage is to remove duplicates before splitting the data (data cleaning). |
| Deep Learning | A subset of Machine Learning that uses neural networks to perform tasks. |
| Defect kit | A collection of samples (or images) with known defects. For example: different types of particulate matter, missing components, damaged primary packaging,… |
| Digital twin | A virtual replicate of a physical object used for simulation. |
| DL | Deep Learning |
| DNN | Deep Neural Network |
| Edge device | In a computer network, an "edge device" is a device that provides an entry point to a network. |
| EMA | European Medicines Agency |
| Ensemble Methods | "ML techniques that combine multiple models to improve the overall predictive performance compared to using a single model" (51) |
| EP | European Pharmacopoeia |
| ETP | Emerging Technology Programme |
| Explainability | "Refers to a representation of the mechanisms underlying AI systems' operation" (51) |
| FDA | Food and Drug Administration (US) |
| Feature | "A pattern in data that can be reduced to a simpler higher-level representation" (52) |
| FMEA | Failure Modes and Effects Analysis |
| FN | False Negative |
| FP | False Positive |
| Frozen model | "A model where all parameters have been finally set, not allowing further adaption to new data" (52) |
| GAMP | Good Automated Manufacturing Practice |
| Generalization | The ability of a model to accurately predict outcomes on new, unseen data by effectively capturing the underlying patterns present in the training dataset. |
| GMLP | Good Machine Learning Practices |
| GMP | Good Manufacturing Practices |
| Grad-CAM | Gradient-weighted Class Activation Mapping. A technique for producing visual explanations for CNN-based models. |
| GUI | Graphical User Interface |
| GxP | Good Practices is an umbrella term, where "x" is a placeholder for a domain. For example: Manufacturing (M), Laboratory (L), Clinical (C), Distribution (D),… |
| GZ | Gray Zone (Knapp methodology) |
| High impact/risk setting | Use cases where errors have large consequences, e.g. related to the primary endpoint in a clinical trial or the physical safety of humans or animals. |
| HMA | Heads of Medicines Agencies. A network of human and veterinary medicines agencies of the European Economic Area (EEA). |
| Hyperparameter | Unlike parameters, hyperparameters are not learned from data. Instead, they are set prior to training of a ML model that governs the learning or model architecture. |
| IaaS | Infrastructure as a Service |
| ICH | International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use |
| IG-NB | Interessensgemeinschaft der Benannten Stellen für Medizinprodukte in Deutschland |
| Inference | Applying a trained model to new, unseen data in order to generate predictions or decisions. |
| INTERN | International Pharmacopoeia (WHO) |
| Interpretability | "Refers to the meaning of AI systems' output in the context of their designed functional purposes" (51) |
| ISPE | International Society for Pharmaceutical Engineering |
| IVDR | In Vitro Diagnostic Regulation (EU) |
| JBIG2 | An image compression standard. |
| JP | Japanese Pharmacopoeia |
| Knapp Methodology | A statistical method to determine the probability of detection in visual inspection of particulate matter. |
| LIME | "Local Interpretable Model-Agnostic Explanations; a technique that approximates any black box machine learning model with a local, interpretable model to explain each individual prediction." (136) |
| MDR | Medical Device Regulation (EU) |
| MHRA | Medicines and Healthcare products Regulatory Agency (UK) |
| ML | Machine learning - an application of AI that enables systems to learn - i.e., models being trained - from data without being explicitly programmed. |
| MNIST database | Modified National Institute of Standards and Technology. A large database of handwritten digits. |
| Model | "Mathematical algorithms with parameters (weights) arranged in an architecture that allows learning of patterns (features) from training data" (52) (136) |
| Model Card | "A structured report of relevant technical characteristics of an AI model and benchmark evaluation results 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" (51) |
| MVI | Manual Visual Inspection |
| NIST | National Institute of Standards and Technology (US) |
| NN | Neural Network |
| OCP | Office of Combination Products (US) |
| OECD | Organisation for Economic Cooperation and Development |

| Acronym / Term | Definition |
|---|---|
| OQ | Operational Qualification |
| Overfitting | “Learning details from training data that cannot be generalised to new data” (52) |
| PaaS | Platform as a Service |
| Parameter | A parameter is a variable that the model learns during training. |
| PCCP | Predetermined Change Control Plan |
| PDA | Parenteral Drug Association |
| PE | Pharmaceutical Engineering (ISPE magazine) |
| Performance Metrics | “In the context of AI quantitative or qualitative measures that can be used to assess the ability of a model to produce the desired output for a given task. The choice of the metrics depends on the specific task and the model objectives” (51) |
| PoD | Probability of Detection |
| PQ | Performance Qualification |
| QIG | Quality Innovation Group (EMA) |
| QMS | Quality Management System |
| ResNet50 | A deep CNN architecture developed by the Microsoft Research Group. |
| RMF | Risk Management Framework |
| RNN | Recurrent Neural Network |
| RQ | Research Question |
| RZ | Reject Zone (Knapp methodology) |
| RZE | Reject Zone Efficiency (units rejected in the RZ / number of units in RZ). A measure used to compare inspection performance. |
| SaaS | Software as a Service |
| SAVI | Semi-Automated Visual Inspection |
| SHAP | Shapley Additive Explanations; an explainable AI (XAI) framework that can provide model-agnostic local explainability for tabular, image, and text datasets |
| SLA | Service Level Agreement |
| SME | Subject Matter Expert |
| Static | “Frozen model : A model where all parameters have been finally set, not allowing further adaption to new data.” (136) |
| Synthetic Data | “Data that have been created artificially (e.g., through statistical modeling, computer simulation) so that new values and/or data elements are generated. Generally, synthetic data are intended to represent the structure, properties and relationships seen in actual patient data, except that they do not contain any real or specific information about individuals” (51) |
| Test dataset | “Data used to evaluate the performance of the AI system, before its deployment. It is expected to be similar to production data, and proper evaluation needs test data to be disjoint from any data used during development” (52) |
| TN | True Negative |
| TP | True Positive |
| Training dataset | “Data used specifically in the context of machine learning: it serves as the raw material from which the machine learning algorithm extracts its model to address the given task” (52) |
| Transformation | Change between different representations of data |
| Transparency | The possibility to fully trace information flow within a ML model |
| Underfitting | “In ML, underfitting happens when a model does not capture the patterns and complexity of the training data, leading to poor performance on both the training and new, unseen data” (51) |
| URS | User Requirements Specification |
| USP | United States Pharmacopoeia |
| V&V | Verification & Validation (ASME) |
| Validation dataset | “Data used by the developer to make or validate some algorithmic choices (hyperparameter search, rule design, etc.)” (52) |
| Variance | The amount by which the performance of a predictive model changes when it is trained on different subsets of the training data. |
| VGG16 | A deep CNN architecture developed by the Visual Geometry Group at the University of Oxford with 16 layers. |
| Visible | Not clearly definable by size but rather through the probability of detection (PoD) that is greater than 70 % - called Reject Zone (RZ) |
| XaaS | As a service model, where “X” is a placeholder – see IaaS, PaaS, SaaS. |
| xAI | Explainable Artificial Intelligence. A field of research that explores methods to counter the “black box” aspect of models. |