IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

THE ROLE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN QA PROCESSES

Sakshi Rajendra Kulkarni[1] , Aniket Sharad salunke[2], Komal Samadhan Zade[3] , Mrunali kadu Ahhirao[4], Prof. Madhuri Ramdas Shirsath[5]

S.V.S Institute of Pharmacy Mungase, Malegaon ,Nashik .

❖ **Abstract**

The term artificial intelligence (AI) describes a machine's capacity to perform tasks that are often accomplished by employing human intelligence, such as learning, asking questions, solving issues, and making judgments. Synthetic intelligence makes use of rule-based structures, neural networks, expert structures, and device learning. AI uses methods and algorithms to organize data, make inferences from laws and patterns, and improve performance over time. The process of software checking out is used to assess and validate a software utility's or product's purported capabilities. Insect protection, lower improvement costs, and enhanced performance are all advantages of trying out. And transportation (CI/CD) pipelines, automation systems gradually lose some of their utility. The process of testing software takes a lot of time, effort, and fatigue. To improve quality and delivery time, automation technologies have been developed to assist in automating certain testing process tasks. Automation solutions increasingly lose some of their utility as pipelines for continuous integration and delivery (CI/CD) are introduced. Because AI is capable of checking code for errors and defects far more quickly than humans and without requiring human intervention, the testing community is looking to AI to fill the gap. Understanding how AI technology affects different STLC jobs or software testing components is the goal of this study. Additionally, the study attempts to identify and elucidate some of the most significant obstacles that software testers encounter when integrating AI into testing. Additionally, the research identifies a number of noteworthy ways AI could advance software

❖ **keywords**

Artificial intelligence, machine learning, Quality Control, Quality assurance,

❖ **Background**

The background of this work is provided in this section, along with an explanation of artificial intelligence (AI), AI quality, and AI quality assurance.

- **AI Quality**

For AI systems to be successful, quality is essential. Numerous research have looked at AI quality challenges from various angles. Four requirement engineering difficulty areas contextual definitions and requirements, data attributes and requirements, performance definition and monitoring, and human factors were identified by Heyn et al [1] as potentially affecting the quality of AI systems. A study on data quality issues in AI systems was carried out by Whang et al [2] They discussed data validation, cleaning, and integration methods and noted the difficulties in each of these domains, such as small, soiled, biased, or contaminated data, and they found numerous practical ways to deal with these issues. One of the most prevalent quality problems during model training, overfitting, was methodically examined by Ying [3], who thoroughly examined its Effective solution.

- **Quality Assurance Properties for AI systems**

QA4AI seeks to guarantee that the AI system satisfies quality standards in a number of areas. This makes it necessary to specify precisely which factors should be taken into account when assessing the caliber of AI systems. Eight QA4AI properties correctness, model relevance, robustness, security, efficiency, fairness, interpretability, and privacy—were outlined by Zhang et al. [4]. Furthermore, numerous research [5, 6, 7, 8,] have noted that many AI system prototypes are challenging to implement. We add deployability in the real world. We also take into account transferability, which has been examined in numerous research [7, 10] and is motivated by the widespread gpractice of using pre-trained models and applying their expertise to different tasks. We add scalability as the eleventh property since many AI systems in the business are made for large user bases, which can lead to issues that compromise the quality of AI systems [11]. The following is a list of definitions for AI quality properties. We summarize the definitions of the final three qualities based on the relevant literature previously mentioned, and we refer to the definitions of the first eight properties from [12].

• **Correctness**: The system's outputs' precision with regard to the tasks it is given. The degree to which an AI model's capabilities match the target data it is meant to analyze in its application environment is known as model relevance.

• **Robustness**: The system's capacity to continue operating steadily in the face of novel, unseen, or loud data as well as environmental changes.

• **Security**: Guarding against internal and external dangers, such as hostile assaults, data breaches, and improper usage of artificial intelligence.

• **Efficiency**: The AI system's capacity to provide results with the least amount of time, energy, or processing power.

• **Fairness**: In AI, fairness refers to the features of AI algorithms that guarantee unbiased judgment, equal treatment of all people, and nondiscrimination based on factors like gender, race, etc

• **Interpretability**: In artificial intelligence, interpretability is a desirable attribute or feature of an algorithm that offers sufficient expressive data to comprehend the algorithm's operation.

• **Privacy**: The system guards against unauthorized access to and use of sensitive and personal data.

• **Deployability**: The simplicity with which an AI system may be included into an already-existing process, environment, or system without compromising its functionality and performance.

• **Transferability**: The system's capacity to apply information acquired in one task or context to another.

• **Scalability**: The system's capacity to sustain or enhance performance as its workload or volume of data grows.

- **Quality Assurance Studies for AI systems**

AI systems differ from typical software systems in many ways, QA4AI is more difficult than QA for regular software systems. The fact that AI systems are data-driven is among the biggest distinctions. The quality of the AI system is directly impacted by the caliber of the data used to train the model, which can also cause the model to behave in an unforeseen way. However, because biases are prevalent in real-world data and many data quality problems are difficult to identify, measuring data quality is extremely challenging.

Numerous investigations have been carried out to tackle the QA4AI issue from various angles. To evaluate DL systems, for instance, Kim et al [14] suggested Surprise Adequacy as an efficient test adequacy criterion. Through retraining, Surprise Adequacy can increase DL systems' resilience to adversarial cases by up to 77.5%, according to an evaluation of the criterion on a variety of DL systems. They do not, however, directly relate to the sector. Even methods presented in carefully chosen research publications about QA for AI systems are difficult for industry practitioners to implement, as Song et al. [15] noted.

QA4AI in the industry is the subject of some studies. For instance, Nahar et al. conducted interviews with 45 individuals from 28 firms, pinpointed three key areas of collaboration and difficulty to illustrate the various organizational styles, and offered suggestions for enhancement. In their research on architecture choices in AI-based system development, Zhang et al. [16] covered the quality factors that developers take into account.

Development of AI-based systems and recognized the prospects and difficulties in this field. The understandability and interpretability of AI models, accuracy and correctness measurements, and dynamic and constantly changing environments are only a few of the difficulties that QA4AI encounters, as noted by Felderer et al. [17].

As stated in Section 1, previous research has examined specific QA4AI facets in the sector. Studies that address participants' perceptions of the significance of QA4AI features, the rationale behind them, as well as the difficulties, related fixes, and industry best practices for each QA4AI trait are, still, lacking. Because of these limitations in the industry's present understanding of QA4AI, we conducted an interview study and a validation survey to learn more about QA4AI and to compile industry experts' best practices.

- **Basic Concepts and Terminologies in Artificial Intelligence And Machine Learning.** [18 ,19]

Offers a thorough dictionary of key terminology, meanings, and examples of each. It serves as a foundational resource for defining terms and acronyms that are often used in the AI-ML community and is connected to important instances in the fields of pathology and medicine. To give readers a firm grasp of the essential terms and ideas that support AI-ML applications, the glossary covers a wide variety of topics, from generative and nongenerative AI to different ML statistical measures.

Data Types for Machine Learning and Artificial Intelligence In the field of AI-ML, data types are crucial since they have a big impact on the methods and strategies employed. According to a summary, the three primary data kinds we come across in medicine are images (including video), text (including voice), and numerical data. Every type of data necessitates specific preparation tasks and is often handled by distinct AI-ML frameworks. Computer vision applications are commonly used for image data, such as the analysis of radiologic pictures (frequently in DICOM format) or WSIs in pathology (now with multiple proprietary file formats).

In order to eventually enable the categorization of such images, machine learning techniques like convolutional neural networks are commonly used for such tasks (e.g., Distinguishing between Cancer and Normal Tissue in Pathology Images). On the other hand, natural language processing (NLP) methods are frequently used to process textual data. These days, these methods mostly use transformer-based architectures to analyze and produce text responses that mimic natural human language, such as generative pretrained transformers (like OpenAI's GPT4 or Meta's Llama3). Lastly, a popular data type used in electronic health record settings is tabular numerical data, which can comprise continuous and/or discrete values and can be used for a variety of predictive modeling applications.

- **Types of Machine Learning.** [20, 21]

The study of learning algorithms is known as machine learning (ML). Learning is the process by which a computer program's performance on a given task, as determined by a performance metric, gets better with practice. The trained computer software is referred to as a model. Therefore, machine learning (ML) models are statistical methods and computational algorithms that allow computers to learn and make judgments or predictions without explicit programming. These models are founded on the ideas of pattern recognition and data-driven learning. Supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning are the four categories into which machine learning can be separated.

1. **Supervised Learning**:

Using labeled data for training, the model gains the ability to predict or classify information. The matching output labels are linked with the input data (features):

• **Classification**: The algorithm learns to predict discrete class labels for input data in this task, such as whether a manufactured product is good or bad. The program gains the ability to predict continuous numerical values through regression. Predicting a production line's energy consumption is one example of this.

2. **Unsupervised Learning:**

In this machine learning technique, a model discovers relationships, patterns, or structures in data without direct assistance from labeled samples. In order to facilitate tasks like clustering, dimensionality reduction, and anomaly detection, it seeks to identify innate patterns or groupings in the data. Unsupervised learning is appropriate for examining and comprehending unlabeled data since, in contrast to supervised learning, it lacks known target labels:

• **Clustering**: Based on patterns or resemblances in the input data, the algorithm clusters related data points together. This includes, for instance, putting machines together that have comparable performance traits.

• **Dimensionality Reduction**: The algorithm in this challenge minimizes the quantity of input features while maintaining crucial data. It can be used as a preprocessing step for other machine learning applications and helps see and comprehend high-dimensional data.

• **Anomaly Detection (AD):** This method involves teaching the algorithm to recognize typical patterns in the data and spotting any data points that drastically differ from those patterns, which frequently point to anomalies or outliers. For instance, this technique can be used to track process machinery signals and detect wear by looking for unusual signal behavior.

• **Autoencoder (AE):** Commonly used for dimensionality reduction and unsupervised learning, an AE is a type of neural network that learns to encode and decode data. It is composed of a decoder network that reconstructs the original data from the latent space and an encoder network that compresses the input data into a lower-dimensional representation (latent space). The objective of an AE is to minimize the reconstruction error, which forces the model to learn a compressed representation that captures the most significant features of the input data.

3. **Semisupervised Learning:** In this learning paradigm, training is conducted using both labeled and unlabeled data. To boost the model's performance, it makes use of a little amount of labeled data in conjunction with a greater amount of unlabeled data. When categorizing data is costly or time-consuming, it can be helpful.

4. **Reinforcement learning :** Is a learning process in which an agent interacts with its surroundings. In order to optimize a reward signal, the agent learns how to behave in its surroundings. The agent gains the ability to make the best choices under various environmental conditions in order to accomplish long-term objectives through a process of trial and error. The primary applications of reinforcement learning are in multistep decision-making tasks including visual navigation, robotics, and video games**.**

❖ Visual Quality Assurance (VQA) [20, 21,22,23]

Is used to describe the procedure of making sure that the visual output such as pictures, movies, and other graphical components meets the necessary requirements for correctness, consistency, and quality. To ensure that the visual depiction of a product or service meets consumer satisfaction and expectations, this process is crucial in a variety of industries, including software development, design, filmmaking, marketing, and manufacturing.

1. **Types of Visual Quality Assurance :**

• **Image Quality Testing:** This entails evaluating an image's quality according to factors such as brightness, contrast, resolution, sharpness, color correctness, and noise levels. These elements guarantee sharp, lifelike, and high-definition visuals.

• **Video Quality Testing:** This involves examining for problems such as color grading, resolution, frame rate consistency, video compression artifacts, and audio-visual synchronization. Visual consistency and fluid playback are guaranteed by video quality assurance.

• **Design Consistency Testing:** This makes ensuring that all design components including layouts, styles, colors, fonts, and logos—follow the established brand guidelines across all media or platforms.

• **Rendering and Performance Testing:** This entails examining the way graphics are rendered and making sure that images work properly on various devices when using 3D rendering in animations or games.

2. **Visual QA Processes :**

• **Automated Testing**: This entails automating the visual testing procedure with software tools. Artificial intelligence (AI) is used by programs such as Applitools to identify visual irregularities and discrepancies between predicted and actual visual output.

• **Manual Testing:** A group of QA testers examines visual output by hand, contrasting it with design criteria and anticipated results. Although this procedure can be time-consuming, it guarantees a greater degree of testing detail.

• **Regression Testing**: Regression testing makes sure that visual components are not impacted by system or program updates or modifications. For example, it's crucial to make sure that prior visual components stay consistent when a new feature is implemented.

• **Cross-Platform Testing:** To guarantee uniformity and usability across all platforms, visual quality assurance also include testing how images show up on various screens, devices, browsers, and operating systems.

3. **Visual Quality Metrics :**

• **Clarity and Resolution:** Verifies if the image or video resolution is suitable for the platform or gadget. Clearer images result from higher resolutions.

• **Color Accuracy:** Verifies that hues and distortions are not present in the reproduction of colors. This is essential for maintaining design coherence, particularly for components linked to a brand.

• **Sharpness and Focus:** Unless purposefully blurred for artistic reasons, this ensures that photos or videos are sharp and focused.

• **Artifacts and Noise:** Visual QA makes sure that the photos or videos are free of undesired artifacts or noise, which can lower visual quality, particularly when compressed.

• **Frame Rate:** To avoid stuttering or visual glitches, be sure that the frame rate (frames per second) of a video is in line with the intended playback speed.

4. **Tools Used in Visual QA :**

• **Applitools Eyes:** An automated visual testing tool that's especially helpful for finding visual errors in mobile and online apps. It compares expected and actual visual outputs using artificial intelligence.

• **Sikuli:** A visual automation tool that recognizes and automates user interface testing through image recognition.

• **Sauce Labs or Browser Stack:** Used for cross-browser testing to make sure images work properly on a range of devices and browsers.

• **Blaze Meter:** This tool is used to evaluate load and performance, particularly for applications that contain visual components like pictures and videos.

5. **Common Visual QA Issues :**

• **Inconsistent Branding:** The user experience and brand identification may be impacted by visual discrepancies between platforms or updates, such as mismatched logos, colors, or font.

• **Resolution Issues:** If images or videos are not appropriately adapted for various screen widths, they may appear pixelated or blurry.

• **Problems with Font Rendering:** Fonts may render differently on various hardware or operating systems, resulting in visual irregularities.

• **Color Variance:** Due to differing color profiles or lighting, colors may appear differently on different screens, which may have an impact on how the viewer perceives the image.

• **Misalignment:** When text, graphics, and user interface elements are not aligned properly, it can create visual clutter and give the product an amateurish appearance.

6. **Why Visual Quality Assurance Matters :**

• **Customer Experience:** Unsatisfactory visual quality can impact user retention and product impression by causing unhappiness.

• **Brand Consistency:** Visual quality assurance makes sure that the brand's image is consistently displayed on all platforms, gadgets, and formats, which fosters professionalism and confidence. Improved Accessibility: Thorough testing guarantees that visual components are usable by those with visual impairments, which promotes inclusion.

• **Error Prevention:** Visual QA assists in identifying little mistakes that, if ignored, could develop into more serious and expensive problems.

7. **Challenges in Visual QA**

• **Expensive Manual Testing:** Manual visual testing necessitates skilled testers and can be time-consuming. Labor costs can be high, particularly when testing intricate images.

• **Automation Challenge:** Although certain testing may be automated using tools like Applitools, it can be difficult to account for every visual scenario, especially when working with dynamic content or unique UI elements.

• **Device Fragmentation:** It might be challenging to maintain visual consistency across several devices with differing sizes, orientations, and resolutions; this calls for thorough testing on various platforms.

• **User Subjectivity:** Various testers may have differing views on what qualifies as acceptable visual quality, making visual quality occasionally subjective.

- **Visual Quality Control (VQC).** [24 -39]

- **Key Methods and Techniques in Pharmacy QC**

  - **High-Performance Liquid Chromatography, or HPLC :** Is frequently used to ascertain the purity and chemical makeup of medications.

- **Spectroscopy :** A method for analyzing active substances quantitatively.

  - **Dissolution testing:** Verifies that the body's drug release rate corresponds to the anticipated performance.

- **Regulatory Standards :** Strict rules and regulations govern pharmaceutical quality control. Among the important regulatory agencies and standards are:

  • **Good Manufacturing Practices (GMP):** A framework for guaranteeing that goods are continuously manufactured and managed in compliance with quality requirements. GMP addresses raw materials, equipment, training, and cleanliness, among other areas of production.

  • **United States Pharmacopeia (USP):** A non-profit that establishes guidelines for the strength, consistency, purity, and quality of medications and their constituents.

  • **European Medicines Agency (EMA):** Offers scientific recommendations for the effectiveness, safety, and quality of medications sold in the EU.

  • **World Health Organization (WHO):** Provides global standards to guarantee pharmaceutical products' efficacy, safety, and quality.

  • **ISO 9001:** While not exclusive to the pharmaceutical industry, ISO 9001 is a popular standard for quality management systems that pharmaceutical businesses can implement to guarantee quality control procedures.

- **Pharmacy Quality Control (QC) :**

In the pharmaceutical sector, pharmacy quality control, or QC, is essential to ensuring that medications fulfill the strictest requirements for efficacy, safety, and quality. Verifying that pharmaceutical products are free of impurities, appropriately dosed, and satisfy consumer and regulatory requirements is the aim of quality control. This is a thorough rundown of quality control in the pharmaceutical sector:

- **The Value of Quality Control in Pharmacy :**

Pharmacy quality control makes sure that goods are produced and maintained in accordance with established guidelines, ensuring patient safety and efficacy. Poor quality control can have detrimental impacts, inefficient therapies, and even deadly outcomes.

A) Essential Elements of Pharmacy Quality Assurance : A Testing of Raw Materials Testing of raw materials, including excipients (inactive components) and active pharmaceutical ingredients (APIs), is the first step in quality control. This guarantees the proper quality of the materials used in the production of drugs.

Tests carried out:

1. API identity and purity
2. Ingredient potency
3. Profiling of impurities
4. Microbial contamination

B) In-Process Control

In order to guarantee constant quality throughout production, quality control (QC) during production makes sure that variables like mixing times, temperatures, and humidity are managed. To keep an eye on the process's uniformity, in-process testing are conducted at different points in time.

Tests carried out:

1. Visual examinations (for uniformity and color)
2. Tests for pH, viscosity, and dissolution
3. Monitoring of humidity and temperature
4. Variability in weight and consistency in composition

C) Finished Product Testing

Before the product is distributed, testing ensures that it satisfies the necessary quality standards after manufacturing.

Techniques used :

1. High-Performance Liquid Chromatography, or HPLC
2. Spectrophotometry in the UV
3. GC, or gas chromatography

D) Microbiological Testing

Examination of Microbes For sterile products (like injectables) or those used in high-risk applications, microbiological testing is particularly crucial. It guarantees that dangerous germs including bacteria, fungus, and endotoxins are absent from medicinal items.

Tests performed

1. Testing for sterility (injectable items)
2. Testing for endotoxins
3. Testing for bioburden
4. Testing for microbiological limits in non-sterile products

E) Stability Testing

Testing for Stability To find out how long a medication will remain potent, safe, and effective under different environmental circumstances (such as temperature, humidity, and light exposure), stability testing is done. It includes real-time stability testing as well as expedited stability testing (under inflated circumstances).

Tests performed :

1. Studies of long-term stability (under suggested storage conditions)
2. Stability studies conducted more quickly (at higher temperatures)
3. Evaluation of degradation products in relation to time.

F) Packaging and Labeling Control

To avoid mistakes in medicine dispensing, QC also makes sure that labels and packaging are completed appropriately. To guarantee correct administration, the labeling must adhere to legal requirements, and the packaging must shield the medication from environmental influences.

Tests performed

1. Packaging materials' integrity
2. Accurate labeling (proper drug name, dosage, expiration date, etc.)
3. Adherence to rules (such as braille labeling for the blind and visually impaired)

G) Documentation and Record Keeping

Record-keeping and Documentation A key component of quality control is accurate documentation and record-keeping. To comply with regulatory bodies, all test, inspection, and process results must be carefully documented and kept on file.

Tests performed :

1. Batch production record
2. Test outcomes from final product testing and in-process control
3. Quality failure investigations (also known as Corrective and Preventive Actions, or CAPA)

- **Pharmaceutical quality control uses various techniques to ensure product quality:**

To guarantee product quality, pharmaceutical quality control employs a number of methods:

1. High-Performance Liquid Chromatography (HPLC): This technique is frequently used to gauge the levels of contaminants, degradation products, and active substances.

2. UV-Visible Spectrophotometry: This technique measures a drug sample's absorbance or reflectance, which is correlated with concentration.

3. Gas Chromatography (GC): Used to determine purity and for volatile compounds like solvents.

4. Dissolution Testing: To guarantee efficient absorption, a drug's rate of dissolution in the body is measured.

5. Microbial Testing: Guarantees product sterility and microbiological safety, particularly for injections and sterile medications.

- **Challenges in Pharmacy Quality Control**

Problems including contamination, human error, and non-compliance with regulatory standards can still occur even with strict quality control procedures. To overcome these obstacles, pharmaceutical companies must consistently make investments in technology advancements, process optimization, and training.

- **Artificial intelligence role in Quality control :**

The role of Quality Control (QC) in AI is critical in ensuring that AI models and systems are effective, reliable, ethical, and fair. This involves overseeing the entire AI lifecycle, from the data collection stage to post-deployment monitoring. Below are the key elements of quality control in AI, with references to research and guidelines:

• Data Quality and Preprocessing:

Making sure the training data is correct, hygienic, and representative is the first step in quality control in AI. Poor data quality has been found to have a direct impact on model accuracy, resulting in bias and inaccurate predictions (Hecht, 2020). Preprocessing methods including feature engineering, data cleansing, and dealing with missing data are essential.

• Model Testing and Evaluation: A number of performance indicators, including F1 score, recall, accuracy, and precision, are used to assess AI models. Models must be validated in order to determine how effectively they generalize to fresh, untested data. Cross-validation, testing, and A/B testing are frequently used for this (Hastie et al., 2009). Thorough assessment aids in guaranteeing that the model operates reliably in a variety of situations.

• Testing for bias and fairness: AI algorithms may inadvertently pick up biased tendencies from training data. An essential component of quality control is identifying and reducing bias. The problem of racial bias in risk assessment algorithms was brought to light by research by Angwin et al. (2016), highlighting the necessity of fairness in AI systems. Methods for identifying bias, include adversarial testing and fairness-aware learning.

• Performance Monitoring After Deployment: To track AI models' performance and make sure they adjust to changes in real-world data, ongoing monitoring is required after they are deployed. This involves monitoring indicators such as drift detection, which indicates when the model's performance gradually declines. According to studies, if the system is not actively maintained, performance may gradually deteriorate (Moreno et al., 2020).

• Governance and Ethical Compliance: A crucial component of quality control in AI is ethical considerations, particularly with regard to privacy, accountability, and transparency. The necessity of making sure AI systems respect human rights and are in line with society values is emphasized in the European Commission's AI ethics guidelines. The QC process includes ethical audits, transparent documentation, and compliance with regulations such as GDPR.

Reference :

1. Hans-Martin Heyn, Eric Knauss, Amna Pir Muhammad, Olof Eriksson, Jennifer Linder, Padmini Subbiah, Shameer Kumar Pradhan, and Sagar Tungal. 2021. Requirement Engineering Challenges for AI-intense Systems Development. In 2021 IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI (WAIN). 89–96. https: //doi.org/10.1109/WAIN52551.2021.00020
2. Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. 2022. Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective. arXiv:2112.06409 [cs.LG]
3. Xue Ying. 2019. An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series 1168 (02 2019), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022
4. Jie M. Zhang, Mark Harman, Lei Ma, and Yang Liu. 2019. Machine Learning Testing: Survey, Landscapes and Horizons. arXiv:1906.10742 [cs.LG]
5. Henrik Hayman, Hendrik Mende, Maik Frye, and Robert H. Schmitt. 2023. Assessment Framework for Deployability of Machine Learning Models in Production. Procedia CIRP 118 (2023), 32–37. https://doi.org/10.1016/j.procir.2023.06.007 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering.
6. Bret Nestor, Matthew BA McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C Hughes, Anna Goldenberg, and Marzyeh Ghassemi. 2019. Feature robustness in non-stationary health records: caveats to deployable model performance in common clinical machine learning tasks. In Machine Learning for Healthcare Conference. PMLR, 381–405.

7. Andrei Paleyes, Raoul-Gabriel Urma, and Neil D. Lawrence. 2022. Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Comput. Surv. 55, 6, Article 114 (dec 2022), 29 pages. https://doi.org/10.1145/3533378
8. Riccardo Perego, Antonio Candelieri, Francesco Archetti, and Danilo Pau. 2020. Tuning deep neural network's hyperparameters constrained to deployability on tiny systems. In International Conference on Artificial Neural Networks. Springer, 92–103.
9. Batia Mishan Wiesenfeld, Yin Aphinyanaphongs, and Oded Nov. 2022. AI model transferability in healthcare: a sociotechnical perspective. Nature Machine Intelligence 4, 10 (2022), 807–809.
10. Mohammad Riyaz Belgaum, Zainab Alansari, Shahrulniza Musa, Muhammad Mansoor Alam, and MS Mazliham. 2021. Role of artificial intelligence in cloud computing, IoT and SDN: Reliability and scalability issues. International Journal of Electrical and Computer Engineering 11, 5 (2021), 4458.
11. Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv preprint arXiv:2006.11371 (2020).
12. Jinhan Kim, Robert Feldt, and Shin Yoo . 2019. Guiding Deep Learning System Testing Using Surprise Adequacy. In 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE). IEEE. https://doi.org/10.1109/icse.2019.00108
13. Qunying Song, Markus Borg, Emelie Engström, Håkan Ardö, and Sergio Rico. 2022. Exploring ML testing in practice – Lessons learned from an interactive rapid review with Axis Communications. arXiv:2203.16225 [cs.SE]
14. Beiqi Zhang, Tianyang Liu, Peng Liang, Chong Wang, Mojtaba Shahin, and Jiaxin Yu. 2022. Architecture Decisions in AI-based Systems Development: An Empirical Study. arXiv:2212.13866 [cs.SE]
15. Michael Felderer and Rudolf Ramler. 2021. Quality Assurance for AI-Based Systems: Overview and Challenges (Introduction to Interactive Session). In Software Quality: Future Perspectives on Software Engineering Quality. Springer International Publishing, 33–42. https://doi.org/10.1007/978-3-030-65854-0_3
16. Maggipinto, M.; Beghi, A.; Susto, G.A. A Deep Convolutional Autoencoder-Based Approach for Anomaly Detection with Industrial, Non-Images, 2-Dimensional Data: A Semiconductor Manufacturing Case Study. IEEE Trans. Autom. Sci. Eng. 2022, 19, 1477–1490. [CrossRef]
17. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. Int. J. Data Min. Knowl. Manag. Process 2015, 5, 1–11. [CrossRef].
18. Albahra S, Gorbett T, Robertson S, et al. Artificial intelligence and machine Learning overview in Pathology & Laboratory Medicine: a general review of Data preprocessing and basic supervised concepts. Semin Diagn Pathol. 2023;40(2):71e87. https://doi.org/10.1053/j.semdp.2023.02.002
19. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and Machine learning in pathology: the present landscape of supervised methods. Acad Pathol. 2019;6:2374289519873088. https://doi.org/10.1177/2374289519873088.
20. Applitools. (2021). Visual Testing with AI – Applitools Eyes. Retrieved from https://applitools.com
21. Upton, J. (2019). Visual Quality Assurance: How to Do It Right. Smashing Magazine. Retrieved from https://www.smashingmagazine.com
22. Zyskowski, K. (2020). The Importance of Visual Quality Assurance in Web Development. Medium. Retrieved from https://medium.com
23. "SikuliX." (2022). SikuliX – Visual Automation Tool. Retrieved from https://sikulix.github.io
24. USP 41-NF 36 : This is the official compendium for drugs, biological products, and dietary supplements in the United States. It provides detailed guidelines and test methods for ensuring drug quality.
25. FDA's Current Good Manufacturing Practices (cGMP) : The U.S. Food and Drug Administration (FDA) sets the standards for pharmaceutical manufacturers to follow to ensure their products are safe, effective, and high quality. (21 CFR Part 211)
26. USP 41-NF 36: Provides methods and criteria for testing raw materials in pharmaceutical manufacturing.

27. FDA's cGMP (Current Good Manufacturing Practices): Guidelines for in-process control, ensuring that pharmaceutical production meets high-quality standards (21 CFR Part 211).
28. USP 41-NF 36 : Specifies testing methods for finished products, including dissolution and potency tests.
29. USP <61> Microbial Limit Tests
30. USP <71> Sterility Tests
31. International Council for Harmonisation (ICH) Guidelines Q1A (R2) : Provides guidelines for stability testing of new drugs
32. FDA Drug Approval Process and Labeling Requirements : Includes regulations on pharmaceutical packaging and labeling.
33. 21 CFR Part 211 : FDA regulations requiring thorough documentation for all QC activities.
34. Hecht, M. (2020). Ensuring Data Quality in AI*. Journal of Data Science & Technology, 12(3), 215-227.
35. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
36. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica.
37. Moreno, M., García, S., & Sánchez, J. (2020). *Quality Control for Artificial Intelligence Systems in Real-Time Applications*. AI & Systems Journal, 16(4), 405-418.
38. Moreno, M., García, S., & Sánchez, J. (2020). *Quality Control for Artificial Intelligence Systems in Real-Time Applications*. AI & Systems Journal, 16(4), 405-418.
39. European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on AI.