AI/ML in Life Sciences

A Summary About the Most Important Technologies for Validation Professionals

DOCUMENT REVIEW HISTORY		
Version	***Change***	***Issue Date***
1	Initial Version	21/APR/2025
2	Including GAMP® AI Guide	05/NOV/2025

Table of Contents

1 INTRODUCTION

In the fast-changing world of Artificial Intelligence (AI) and Machine Learning (ML), it can be tough to keep up with all the different techniques and how they are used in Life Sciences. Often, these techniques overlap and connect in many ways.

To make things easier to understand, we have organized a list of key AI and ML technologies and their uses in the pharma, biotech, and medical device sciences fields. This structured approach will help you understand the framework of AI/ML.

We cover topics like machine learning, deep learning, natural language processing, and generative AI, showing how these tools are changing medical research, diagnostics, and treatment planning. Whether you are an expert or just starting, this guide will help you understand AI and ML in Life Sciences, providing insights and examples.

If you have any questions or need more help, please contact us at support.academy@fivevalidation.com or contact@fivevalidation.com

2 INTRODUCING FIVE VALIDATION

FIVE Validation is the company behind GO!FIVE®, a cloud-based digital validation platform that accelerates regulatory compliance by up to 7x using agile methodologies. With over 70 ready-to-use templates, GO!FIVE® simplifies the validation of AI/ML solutions, traditional GxP systems, IT and OT infrastructure, equipment, and processes, turning compliance into a strategic asset.

FIVE delivers digital validation services across the entire product lifecycle and supports more than 200 Life Sciences clients globally. With offices in Brazil, the Netherlands, and the United States, the company is certified to ISO 9001:2015 and ISO 27001:2022 standards.

3 HOW TO VALIDATE AI/ML

If you're looking to learn how to validate these technologies in practice, check out our blogs on

How to Validate AI/ML and AI Data Governance.

4 REGULATIONS AND GUIDELINES TO KNOW

Understanding the regulatory landscape is essential when working with AI and ML in Life Sciences. These technologies must align with local and international expectations for safety, efficacy, transparency, and accountability. That’s why it’s important to stay up to date with the main regulations and guidelines currently shaping this field across the globe.

Here are some of the key regulatory references you should be aware of:

- ISPE GAMP®* Guide: Artificial Intelligence, JUL 2025
- FDA AI/ML-Based Software as a Medical Device Action Plan, JAN 2021
- FDA, MHRA, Health Canada: Good Machine Learning Practice for Medical Device Development, OCT 2021
- FDA Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions, DEC 2024
- FDA, Draft: Artificial Intelligence–Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations, JAN 2025
- AI European Act, JUN 2024
- EMA Reflection Paper on the Use of AI in the Product Lifecycle, SEP 2024
- MHRA Software and AI as a Medical Device: Change Programme – Roadmap, JUN 2023
- Health Canada: Guidance Document: Software as a Medical Device (SaMD): Definition and Classification, DEC/2019
- ISO/IEC TR 29119-11:2020 Software and systems engineering – Software testing Part 11: Guidelines on the testing of AI-based systems (free of cost)
- ISO/IEC TR 24027:2021 Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI-aided decision making
- ISO/IEC 22989:2022 Information Technology – Artificial Intelligence - concepts and Terminology (free of cost)
- ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- ISO/IEC 42001:2023 Information technology – Artificial Intelligence – Management system
- AMI TIR34971:2023 Application of ISO 149071 to Machine Learning in Artificial Intelligence – Guide

These documents provide the foundation for responsible AI/ML adoption in regulated environments and are essential reading for anyone working with digital health technologies in Life Sciences.

* GAMP® is a brand that has its intellectual rights reserved by ISPE®. Available for purchase at https://ispe.org/.

5 AI TECHNIQUES

5.1 Machine Learning

It is a type of AI where systems learn from data and improve over time without explicit programming.

5.1.1 Types of Learning

- **Supervised learning:** Learning from labeled data to predict future outcomes (e.g., classification, regression).
- **Unsupervised learning:** Finding hidden patterns in unlabeled data (e.g., clustering, dimensionality reduction).
- **Reinforcement learning:** Learning through trial and error by receiving rewards or penalties to optimize decisions.

5.1.2 Neural Networks

Neural networks are a type of machine learning model inspired by the human brain. They consist of layers of interconnected nodes (neurons) that process data and learn to recognize patterns for tasks like classification and prediction.

- **CNNs (Convolutional Neural Networks)** are neural networks designed to interpret spatial data, such as images, by recognizing local patterns through convolutional and pooling layers.
- **RNNs (Recurrent Neural Networks)** are neural networks specialized in sequential data, such as text and speech, using memory of previous inputs to capture temporal dependencies across the sequence.

5.1.2.1 Neural Networks Applications

- **Medical Image Analysis**: Neural networks are used to analyze and classify medical images (e.g., X-rays, MRI, and CT scans) to identify diseases such as cancer, brain injuries, and bone fractures, improving diagnostic accuracy.

- **Disease Outbreak Prediction**: Neural networks model epidemiological data to predict disease outbreaks, aiding in resource allocation and emergency response management.

- **Genomic Analysis**: Neural networks identify patterns in DNA data, facilitating the discovery of genetic mutations linked to rare diseases or predispositions.

- **Treatment and Drug Response Prediction**: Neural networks analyze clinical data to predict how patients will respond to specific treatments or medications, enabling personalized therapies and optimizing treatment plans based on patient-specific data, including genetic profiles and medical history.

- **Automated Clinical Triage**: Neural networks help classify and prioritize patients in hospitals based on the severity of symptoms, enhancing healthcare efficiency.

- **Predicting Patient Outcomes**: Estimating outcomes such as recovery time, complications, or survival rates based on medical records and patient-specific data.

- **Predicting Drug Side Effects**: Identifying unknown side effects of drugs by analyzing patient data and detecting adverse reactions.

- **Predicting Disease Progression**: Forecasting disease progression based on patient data, including symptoms, biomarkers, and medical history.

- **Biomarker and Cell Type Classification**: Categorizing biomarkers that indicate the presence or severity of diseases (e.g., cancer markers) and identifying different cell types (e.g., cancerous vs. healthy cells) from microscopic images or genetic data to aid in research and diagnostics.

5.1.3 Predictive Analytics

It involves using statistical and machine learning techniques to analyze historical data and forecast future trends or outcomes. It aims to identify patterns and relationships in the data to make informed predictions.

5.1.3.1 Predictive Analytics Applications

- **Predicting Disease Progression**: Forecasting disease progression based on patient data, including symptoms and biomarkers.

- **Adaptive Clinical Trials**: Dynamically modifying treatment regimens during trials based on patient responses and ongoing results.

- **Optimizing Drug Discovery Process**: Identifying optimal drug candidates by efficiently navigating chemical spaces.

- **Personalized Treatment Plans for Chronic Diseases**: Optimizing treatment strategies for chronic diseases based on patient responses and predictive outcomes.
- **Adaptive Radiation Therapy for Cancer**: Tailoring radiation doses based on tumor response to minimize damage to healthy tissue.

- **Personalized Medicine**: Utilizing generative AI and predictive models to simulate patient responses, allowing for individualized treatment plans across various medical conditions.

- **Pharmacokinetics**: Modeling drug absorption, distribution, metabolism, and excretion in the body, considering patient-specific factors such as age and organ function.

- **Predicting Patient Outcomes**: Estimating the likelihood of outcomes like hospitalization or recovery based on patient data, including symptoms and treatment history.

5.1.4 Transfer Learning

Transfer Learning is a machine learning technique where a model developed for one task is adapted and applied to a different but related task.

5.1.4.1 Transfer Learning Applications

- **Tumor Recognition**: Pre-trained models for medical image recognition are fine-tuned to identify specific types of tumors in various cancers and automatically analyze pathology slides to detect cancerous cells or tissues.

- **Rare Disease Classification**: Transfer learning allows AI models trained on common diseases to recognize patterns in rare diseases, even with limited data.

- **Diagnostic Prediction Models**: Systems leverage data from other clinical contexts to improve the accuracy of condition or diagnosis predictions.

- **Drug Development**: Pre-trained models on large pharmacological datasets are used to predict drug interactions in new drug combinations.

- **Biomarker Identification**: Transfer learning assists in identifying new biomarkers in genomic analyses using data from other biomedical research fields, and in genomic data analysis, it identifies relationships between genetic markers and specific traits.

5.1.5 Generalized Linear Model (GLM)

A Generalized Linear Model (GLM) extends traditional linear regression to handle various types of data, such as binary or count outcomes. It uses a link function to connect a linear combination of input variables to the mean of the target variable, allowing the model to fit different data distributions, making it more flexible for diverse scenarios.

5.1.5.1 GLM Applications

- **Logistic Regression:** A type of GLM used for binary outcomes, such as predicting whether an email is spam or not.
- **Poisson Regression:** Used to model count data, such as the number of customer calls in an hour.
- **Gamma Regression:** Models positive continuous data with skewed distributions, commonly used in insurance claims data.
- **Interpretable Models**: In machine learning workflows, GLMs are foundational models that offer clear insights into the relationship between input variables and outcomes, especially for structured, tabular data.
- **NLP Application:** GLMs, like logistic regression, can be used in Natural Language Processing (NLP) tasks such as text classification.

5.2 Deep Learning

Deep Learning is a subset of machine learning that uses neural networks with many layers (deep networks) to analyze complex patterns in large datasets, often achieving high accuracy in tasks like image recognition, natural language processing, and speech recognition.

5.2.1 Deep Learning Applications

- **Medical Image Analysis**: Classifying medical images (e.g., X-rays, MRIs, CT scans) and automatically analyzing pathology slides to detect diseases such as cancer, pneumonia, fractures, or other abnormalities.
- **Anomaly Detection in Medical Imaging**: Identifying unusual patterns in medical images (CT scans, MRIs) that may indicate tumors or other abnormalities.

- **Image-Guided Surgery**: Assisting surgeons during operations by identifying anatomical structures in real-time medical images.

- **Protein Structure Prediction**: Predicting the 3D structure of proteins from amino acid sequences to understand their function.

- **Patient Risk Stratification**: Classifying patients based on risk levels for surgical complications or developing conditions like cardiovascular disease or stroke, using pre-operative data and lifestyle factors.

- **Predicting Medical Device Performance**: Estimating the performance or lifespan of medical devices (e.g., implants, pacemakers) based on patient demographics and usage patterns.

- **Survival Analysis**: Predicting patient survival duration after treatment or diagnosis based on factors such as treatment type, age, and existing health conditions (comorbidities).

- **Modeling Disease Progression**: Forecasting the course of diseases such as diabetes or cancer using clinical variables like blood glucose levels, tumor size, or biomarker levels.

- **Cancer Detection in Medical Imaging**: Using computer vision to analyze images such as mammograms and CT scans to accurately detect abnormalities or tumors.

- **Vital Sign Monitoring**: Computer vision systems monitor patients' vital signs in real-time, including respiratory movements and facial expressions.

- **Ocular Pathology Detection**: Detecting eye diseases like diabetic retinopathy and macular degeneration using retinal images.

- **Assisted Surgery Automation**: Cameras and computer vision algorithms guide robotic surgical procedures, increasing precision and assisting surgeons in complex tasks.

- **Automatic Exam Screening**: Computer vision analyzes images from lab tests, such as biopsy slides, to speed up the diagnostic process.

- **Automated Cell Counting and Classification**: Automatically counting and classifying cells in biological samples.

- **Training Autonomous Medical Robots**: Developing robots for diagnostics, treatment delivery, or surgery assistance.

5.3 Natural Language Processing (NLP)

It is a field of artificial intelligence that enables computers to understand, interpret, and generate human language. It involves tasks like language translation, sentiment analysis, and text generation. NLP relies on large datasets to train models, which learn the structure and nuances of human language through statistical and machine learning techniques.

5.3.1 NLP Applications

- **Electronic Health Record (EHR) Analysis**: NLP processes and extracts relevant information from free-text medical records, aiding in diagnosis and treatment planning.
- **Clinical Chatbots**: NLP systems interact with patients to answer questions about symptoms and assist with appointment scheduling.
- **Patient Narrative Analysis**: NLP processes verbal reports and patient notes, identifying trends in symptoms or treatment responses.
- **Regulatory Document Automation**: NLP is used to generate or review regulatory documents, such as clinical trial submissions and compliance reports.
- **Medical Literature Search**: NLP tools conduct searches in medical databases and scientific literature, speeding up the discovery of relevant information.

5.4 Generative AI

It refers to artificial intelligence systems that create new content, such as text, images, or music, based on patterns learned from existing data. It can generate original outputs that mimic the style or characteristics of the input data.

5.4.1 Generative AI Applications

- **Optimizing Gene Editing with CRISPR**: Optimizing gene-editing protocols using CRISPR to maximize gene correction and minimize off-target effects.
- **Drug Discovery and Design**: Generative AI creates novel drug molecules by generating chemical structures with desired properties like high efficacy and low toxicity.

- **Personalized Medicine**: Generative AI models simulate how different patients might respond to various treatments, assisting doctors in designing personalized treatment plans and predicting drug efficacy based on factors such as age, weight, and genetic characteristics.

- **Medical Report Generation**: Generative AI automates the generation of medical reports, such as radiology reports, based on image analysis or diagnostic data.

- **Optimizing Clinical Trials**: Generative AI simulates patient outcomes to help design better clinical trials and predict potential side effects or efficacy.

- **Medical Image Enhancement and Reconstruction**: Generative AI enhances or reconstructs low-quality medical images, improving diagnostic accuracy.

- **AI-Driven Virtual Assistants for Healthcare**: Generative AI powers virtual assistants that assist patients and healthcare providers by generating responses to medical queries.

5.4.2 LLMs (Large Language Models)

LLMs are a type of AI that uses vast amounts of text data to understand and generate human-like language. They can perform tasks such as translation, text completion, and question answering by learning patterns and relationships in the data they were trained on.

5.4.2.1 LLMs Applications

- **Identifying Hidden Biomarkers in Disease**: Discovering new biomarkers by analyzing large-scale omics data.

- **Discovering Disease Subtypes**: Identifying new subtypes of diseases based on gene expression profiles and clinical outcomes.

- **Drug Discovery and Development**: LLMs process large amounts of scientific literature, patents, and clinical trial data to identify potential drug candidates and mechanisms of action, as well as classify the likelihood of adverse drug reactions based on medical history and genetics.

- **Biomedical Literature Mining**: LLMs help researchers extract key information from vast amounts of published biomedical research.

- **Clinical Trial Management**: LLMs organize, analyze, and interpret extensive text-based data generated by clinical trials, including patient notes and protocols.

- **Personalized Medicine and Genomics**: LLMs assist in analyzing genomic data by interpreting and summarizing complex genetic sequences.

- **AI-Assisted Diagnostics**: LLMs help interpret clinical data such as pathology reports, lab results, and radiology findings to support diagnoses.

- **Clinical Decision Support**: LLMs synthesize vast amounts of medical data to provide recommendations for treatment options or diagnostics.

- **Patient Interaction and Education**: LLMs communicate with patients by answering medical queries, explaining procedures, and providing medication instructions.

5.4.3 Multimodal Foundation Model (MfM)

MfM is a type of generative AI. It represents a model architecture that integrates and processes multiple types of data (e.g., text, images, audio) to generate and understand complex outputs across different modalities. MfMs enable the creation of rich, context-aware content by leveraging diverse data inputs.

5.4.3.1 MfM Applications

- **Integrated Image and Text Analysis**: Multimodal models combine medical image data with patient notes to provide more accurate diagnoses.

- **Clinical Trial Monitoring**: Multimodal models integrate data from various sources (text, images, device data) to monitor real-time results in clinical trials.

- **Genomics and Image Data Fusion**: Models integrate genomic sequencing data with medical images to identify correlations between genetic mutations and clinical visual conditions.

- **Patient Monitoring**: Applications integrate data from biometric sensors, written reports, and visual signals to provide a comprehensive view of a patient's health, including real-time monitoring to recommend interventions based on physiological data.

- **Automated Drug Delivery Systems**: Optimizing drug delivery by adjusting doses in real-time based on physiological signals.

5.4.4 Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) is an AI technique that combines retrieval-based methods with generative models. It enhances generative capabilities by retrieving relevant information from external sources during the generation process, improving the accuracy and relevance of the output.

5.4.4.1 RAG Applications

- **Pharmaceutical Research and Development:** RAG can generate reports or analyses by retrieving information from scientific publications, clinical trial data, and regulatory guidelines, ensuring compliance and alignment with current regulations.

- **Quality Management and Regulatory Compliance:** RAG improves the transparency and traceability of AI-generated outputs by linking them to the sources of information. This is critical for auditing and validating AI systems, meeting the requirements of regulatory bodies.

5.5 Genetic algorithms

They are optimization methods inspired by natural selection. Using selection, crossover, and mutation, they evolve solutions to complex problems, searching for optimal outcomes in large search spaces. By evaluating solutions with data, the algorithm iteratively improves, learning which traits lead to better results over generations.

5.5.1 Genetic algorithms Applications

- **Drug Discovery and Toxicity Prediction**: Predicting drug toxicity or efficacy by analyzing chemical structures and exploring chemical space to identify promising drug candidates.

- **Optimizing Drug Discovery Process**: Utilizing genetic algorithms to efficiently navigate chemical spaces and simulate the evolution of drug molecules for optimal design.

- **Optimization of Bioprocesses**: Enhancing fermentation processes, enzyme activity, and bioreactor conditions for producing pharmaceuticals, biofuels, or industrial enzymes.

- **Genomic Sequence Alignment and Analysis**: Aligning DNA or RNA sequences and analyzing large-scale genomic data to uncover patterns in genetic variations.

- **Optimizing Cancer Treatment**: Developing personalized cancer treatment strategies by modeling and predicting effective therapy combinations based on patient data.

- **Optimization of Diagnostic Tools**: Enhancing medical diagnostic algorithms for image recognition in detecting diseases like cancer or heart disease.

- **Gene Network Modeling**: Modeling gene regulatory networks to explain gene interactions in biological processes.

- **Epidemiological and Clinical Trial Modeling**: Assisting in the modeling of strategies to control infectious diseases and designing adaptive clinical trials based on patient responses.

- **Clustering Patients for Personalized Medicine**: Grouping patients with similar characteristics to create tailored treatment plans.

- **Discovering Biomarker Patterns**: Identifying novel biomarkers indicating disease presence or progression using unstructured clinical or molecular data.

- **Cell Type Identification in Single-Cell RNA Sequencing**: Classifying different cell types within large datasets to measure gene expression profiles.

- **Anomaly Detection in Medical Imaging**: Detecting anomalies in medical images that may indicate diseases such as tumors.

- **Identifying Unknown Disease Subtypes**: Analyzing patient data to find new subtypes of diseases based on gene expression and symptoms.

- **Functional Grouping of Genes or Proteins**: Grouping genes or proteins with similar functions based on expression data or interaction networks.

- **Detecting Drug Side Effects**: Identifying unknown side effects of drugs through patient-reported and clinical trial data.

- **Discovering Novel Microbial Communities**: Classifying and identifying microbial communities in environmental or clinical samples.

- **Gene Therapy and Drug Formulation Optimization**: Optimizing gene therapy protocols and drug formulations to improve effectiveness and stability.