

## OPINION PAPER

# Current Opportunities for the Integration and Use of AI/ML in Clinical Trials: Good Clinical Practice Perspectives

Joseph Geraci\*, Prasanna Rao†, Cheryl Grandinetti‡, Bessi Qorri\*, Patrick Nadolny§, Kassa Ayalew‡, Lisbeth Bregnhøj||, Lindsay Edwards¶, Karen Hofmann\*\*, Sean Khozin††, Nicolas Schaltenbrand‡‡, Torsten Stemmler§§, Alan Yeomans|||, Demetris Zambas¶¶ and Ni A. Khin\*\*\*

The integration of artificial intelligence and machine learning (AI/ML) in clinical trials offers immense potential to reshape drug development and research efficiency. This review explores the multifaceted landscape of AI/ML applications in clinical trials including seven use cases of AI/ML that aim at improving data quality and enhancing patient outcomes and clinical trial successes. We discuss good machine learning practices focusing on clear scope definition, transparent risk-proportionate management, and robust feature engineering. We also highlight privacy-preserving data-sharing techniques like federated learning and the role and the risk of using large language models in patient recruitment, data capture, clinical decision support, patient engagement, and trial design optimization. This review highlights the potential of aiding clinical trials through the responsible use of AI/ML, while recognizing the challenges (e.g., generalizability, transparency, and robustness) as well as ethical considerations including patient safety, privacy, and human rights. It also serves to guide interested parties towards the responsible and effective integration of these technologies with clinical trials.

**Keywords:** Artificial Intelligence (AI); Machine Learning (ML); Clinical Trials, Good Clinical Practice (GCP); Drug Development; Regulatory Considerations

## Introduction

Artificial intelligence (AI) and machine learning (ML) technologies are rapidly transforming the healthcare sector, holding immense potential to reshape the landscape of medical practice and clinical research,

improving both the patient experience and clinical outcomes.<sup>1,2</sup> AI/ML adoption is also catalyzing major shifts in the conduct of biomedical research by facilitating rapid analysis of big data, enabling researchers to gain valuable insights at an unprecedented pace.<sup>3</sup>

The pharmaceutical industry has already begun harnessing the power of AI/ML throughout the drug discovery and development process.<sup>4</sup> This shift presents unique opportunities for the design, conduct, and analysis of more efficient and effective clinical trials. However, as the role and implementation of AI/ML in clinical trials continues to expand and evolve, there arises a need for good clinical practice (GCP) advice and assessments. Such GCP perspectives are essential to evaluate and ensure the quality, integrity, and reliability of AI/ML-driven data submitted to regulatory agencies, crucial for supporting the safety and efficacy of medical products and regulatory decision-making.<sup>5</sup>

In response to these pivotal developments in AI/ML adoption in clinical drug development, the European Medicines Agency's (EMA) GCP Inspectors Working Group hosted two meetings in 2020 and 2021. These meetings were dedicated to exploring the integration of AI/ML in clinical trials. The 2020 meeting delved into the

\* NetraMark Corp, Toronto, ON, CA

† Saama Technologies, Campbell, CA, US

‡ US Food and Drug Administration (USFDA), Center for Drug Evaluation and Research, Silver Spring, MD, US

§ Sanofi S.A., Paris, FR

|| Danish Medicines Agency (DKMA), DK

¶ Relation, London, UK

\*\* Cognizant Technology Solutions, Bethlehem, PA, US

†† PhysionBio LLC, New York, NY, US

‡‡ Wega Informatik AG, Basel, CH

§§ Federal Institutes for Drugs and Medical Devices (BfArM), Bonn, DE

||| Viedoc Technologies AB, Uppsala, SE

¶¶ Pfizer, Inc, New York, NY, US

\*\*\* Neurocrine Biosciences, Inc., San Diego, CA, US

Corresponding author: Cheryl Grandinetti, Pharm.D.  
(Cheryl.Grandinetti@fda.hhs.gov)

characterization of the ideal conditions, prerequisites, and illustrative real-world use cases for AI/ML in biomedical research and therapeutic development.<sup>6</sup> Building upon this foundation, the 2021 meeting explored the nuances of good ML practices while navigating the ethical and regulatory considerations that the integration of AI/ML entails.<sup>7</sup>

We acknowledge that several years have passed since these meetings were held. However, the foundational considerations remain highly relevant in light of the evolving use of AI/ML technologies in clinical trials. This time lag underscores the importance of continuous reflection and underscores the value of publishing periodic summary papers to capture key insights, assess emerging developments, and inform ongoing efforts in responsible AI/ML implementation within clinical trials.

In this paper, we aim to shed light on the critical considerations arising from these meetings, offering a comprehensive overview of the key discussion points, while also considering developments that have since come into view in this quickly evolving field. By distilling the insights and knowledge shared during these gatherings, and through our own practice with AI/ML, we aim to chart a course towards harnessing the full potential of AI/ML technologies in clinical trials while upholding the GCP standard of patient safety and trial integrity to ultimately advance health outcomes for all. It is important to note

that the principles and suggestions presented herein are intended to be high-level rather than prescriptive, recognizing the need for flexibility in application across diverse contexts.

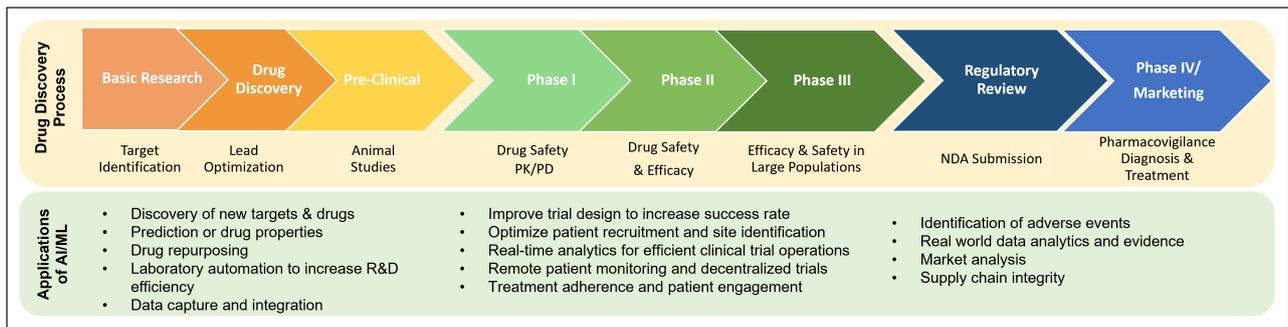
### Overview of AI/ML Approaches and Their Role in Global Clinical Trials

Before providing a comprehensive examination of AI/ML applications in global clinical trials, it is important to delineate the different approaches employed throughout the drug discovery and development continuum. **Table 1** offers an overview of the different types of AI/ML approaches used in drug development, which include supervised learning, unsupervised learning, reinforcement learning, transfer learning, natural language processing (NLP), and generative models.<sup>4</sup> AI/ML technologies have utility across the entire spectrum of drug development, starting with basic research and target identification, novel molecule generation, and extending through to clinical trials and market access (**Figure 1**).<sup>4</sup> Specifically at the clinical trial stage, ML methods are currently making substantial contributions by:

- Evaluating and ranking compounds based on their potential for success.
- Anticipating and mitigating toxicity events.
- Enhancing patient recruitment efforts for Phase II and III trials.

**Table 1:** Types of AI/ML approaches in drug development.

AI/ML Approach	Description	Potential Use in Drug Development	Refs
<b>Supervised Learning</b>	Uses labelled data to discover predictive models	Predicting drug response, placebo response, and adverse events. Also used to discover novel drug targets	11
<b>Unsupervised Learning</b>	Uses unlabelled data to discover patterns or structures in the data	Helping to identify unknown patient heterogeneity, and explore complex biochemical factors driving drug response	11
<b>Semi-Supervised and Self Learning</b>	Combines a small amount of labeled data with a larger pool of unlabeled data to improve predictive accuracy; iteratively labels and trains on unlabeled data, relying on their own predictions to refine learning	Analyzing sparse or incomplete clinical trial data by combining small, labeled datasets with larger unlabeled datasets, uncovering biomarkers, subpopulations, or drug response patterns while reducing the need for extensive labeled data collection	12,13
<b>Reinforcement Learning</b>	Enables machines to learn by interacting with an artificial interactive environment under a reward scenario	Optimizing the selection of molecules for synthesis; designing adaptive clinical trials; discover patient response personas	14
<b>Transfer Learning</b>	Uses knowledge gained from one task to improve the performance of another, often related task	Overcoming limitations of small datasets in drug development; leveraging knowledge from related domains; using pre-trained neural networks for image analysis in medical imaging and drug discovery	15
<b>Natural Language Processing and Large Language Models (LLMs)</b>	Enables computers to process human language and generate human readable explanations	Mining scientific literature, patents, and clinical trial data for relevant insights to aid drug discovery and provide explainability	16
<b>Generative Models</b>	Uses large amounts of data including molecular Statistical Machine Intelligence and Learning Engine (SMILES), patient descriptors, and biochemistry for pre-training	Allowing groups to generate artificial instances of patients to increase sample sizes and better understand patient populations, and it is currently the "go-to" method by which new drugs are discovered through AI	17,18



**Figure 1: Use and application of AI/ML throughout the entire drug development process.** Created and modified with permission from Schaltenbrand 2020.<sup>19</sup>

- Enriching eligibility criteria for improving drug and minimizing placebo response.
- Providing enrichment criteria to de-risk clinical trials.
- Creating predictive models using digital twins.
- Developing biomarkers to enable precision medicine applications.<sup>8</sup>

Recent strides in transformer-based methods at the center of large language models (LLMs) have ushered in novel possibilities.<sup>9,10</sup> LLMs offer significant potential to communicate insights in a clear and accessible manner. However, their effectiveness in clinical trial settings depends on the prior application of advanced machine intelligence methods capable of extracting meaningful, patient-level patterns from complex data. These upstream systems perform the critical task of elevating raw information into structured insights, which LLMs can then contextualize and present to clinical trialists. This process strengthens the connection between scientific discovery and clinical trial leadership by allowing for more informed decision-making.

The optimal integration of AI/ML methods in clinical trials hinges on three pivotal domains:

- 1. Data Assets:** Data assets serve as the raw materials for ML algorithm training, encompassing data generated beyond traditional clinical trials, such as real-world data (RWD) as well as previous trial data that can be leveraged to enrich algorithm training.<sup>11,12</sup> Harnessing this wealth of data through the power of AI/ML empowers researchers to gain deeper insights, optimize patient selection, and ultimately accelerate the development of more effective and personalized treatments.
- 2. Advanced Analytical Software:** The robust management of training datasets and algorithm validation is facilitated by advanced analytical software applications.<sup>13</sup> These tools serve as indispensable facilitators in the AI/ML journey.
- 3. Next-Generation Computational Capabilities:** Next-generation computational capabilities may provide opportunities in AI that are not currently feasible using classical computational methods. These capabilities may accelerate algorithmic processing and may open doors to modelling biological and chemical phenomena at the quantum level. While current

hardware platforms pose limitations in this regard, ongoing efforts are poised to overcome these barriers.

The rapid advancements in graphics processing unit (GPU) technology are enabling the training of massive foundational transformer-based technologies. Pharmaceutical companies are leveraging these innovations by partnering with LLM providers to develop customized AI systems. These transformer-driven technologies, powered by increasingly powerful GPUs, are being applied to critical areas such as drug discovery, biological simulations, and the analysis of Electronic Health Records (EHRs), among other drug development applications.<sup>14,15</sup>

Foundational advances in genomic sequencing, transcriptomics, epigenetics, and mass spectrometry technologies have independently contributed to redefining the classification of advanced malignancies. Building on these developments, AI/ML methodologies are now increasingly being leveraged to enhance and accelerate this progress by integrating and analyzing complex, high-dimensional datasets.<sup>16,17</sup> For example, these technologies have enabled disease classification that transcends traditional distinctions such as anatomical location (e.g., lung vs. colon cancer) or histopathological features (e.g., squamous vs. adenocarcinoma), in favor of more precise molecular characterizations (e.g., gene expression profiles, protein levels, microRNA signatures).<sup>18</sup> In 2017, the US FDA approved an immune checkpoint inhibitor for advanced solid malignancies exhibiting a common molecular phenotype, regardless of the anatomical location or histopathology of the tumors.<sup>19</sup> This has paved the way for precision therapies and clinical trial patient enrichment strategies that optimize risk-benefit profiles by leveraging multiomic data and algorithm-trained data assets. AI/ML technologies have the potential for global clinical trials to advance patient care by using these modalities of data to reveal novel taxonomies of disease through this high-resolution lens. This is helping with the pursuit of elucidating disease mechanisms beyond the classical organ model.

### Exploring the Application of AI/ML in Clinical Trials: Seven Real-World Use Cases

In this section we describe seven use cases that were presented at the AI/ML workshop meetings hosted by EMA's GCP Inspectors Working Group that showcase the

versatile applications of specific AI/ML methods in clinical trials, underscoring their potential:

### **Case 1: Smart Data Query**

Traditionally, data managers have been responsible for identifying discrepancies and generating queries using manual techniques such as spreadsheets. ML introduces what is referred to as a “Smart Data Query” that predicts discrepancies, elucidates the reasons behind them, and auto-generates query text, designed for “human-in-the-loop” validation.<sup>20</sup>

One foundational case that helped shape the development of these newer Smart Data Query approaches involved matching adverse events to concomitant medications, a task that requires a large volume of data for identifying discrepancies and clinical inference to understand the potential relationship between drugs and adverse events. While this use case may now appear dated, it demonstrated how ML could augment a traditional data review process. In this particular case, a combination of semi-supervised learning and clinical inference models were employed to identify discrepancies between adverse events and concomitant medications – a task requiring both data analysis and clinical development. A human-in-the-loop approach enabled data managers to assess the logical coherence of concomitant medications with adverse events. Testing of the model’s accuracy involved comparing historical queries raised by data managers with those generated by the ML model. The results of these unified human/ML models revealed that they achieved an 85% to 90% accuracy range, reducing the time required from data entry to query generation by 50% and significantly streamlining the entire workflow.<sup>20,21</sup> This case laid important groundwork for subsequent advances in Smart Data Query systems, many of which now incorporate similar human-in-the-loop mechanisms and logic-driven inference at more sophisticated levels.

### **Case 2: Addressing Data Attributability Challenges in Wearable Devices Using AI/ML Fingerprinting Techniques**

In recent years, the adoption of digital health technologies, such as wearables, has increased the remote collection of trial endpoint data from study participants. Unlike traditional electronic data capture (EDC) systems that rely on user identification through access controls, wearables present a unique challenge as they often lack the ability to attribute data to the individual wearers. AI/ML methods provided an opportunity to overcome this challenge.<sup>22,23</sup>

In this example, AI/ML was used to create a distinctive data fingerprint, as a digital biomarker, for each user by analyzing raw actigraphy data.<sup>24</sup> This entailed the application of pattern recognition techniques to analyze visual representations of three-minute snippets of actigraphy data graphs. Leveraging this wealth of identifiers, the program created digital fingerprints that achieved high accuracy in matching the data to the respective wearers.<sup>25</sup> This use case serves as an example of the potential of AI/ML to overcome complex data attribution challenges, enabling more reliable and insightful clinical trials.

### **Case 3: Enhancing Protocol Deviation Trending**

Protocol deviation (PD) trending is essential for ensuring patient safety, regulatory compliance, and overall data integrity in clinical trials. While central monitoring of data within EDC systems is a foundational tool, it is insufficient on its own for comprehensive PD detection. Many deviations originate or are documented outside the EDC—often buried in free-text fields within Clinical Trial Management Systems (CTMS), monitoring reports, or site logs and communications. A major challenge arises when deviations are logged in open-text fields without predefined categories, leading to a large proportion being labeled as “other” or “non-classified”. This limits visibility and impairs the ability to trend meaningful patterns across sites or studies. As a result, the PD trending process often becomes burdensome and time intensive. LLMs offer a transformative solution by enabling the efficient classification of PDs into predefined categories using advanced NLP techniques, thereby enhancing study oversight and streamlining operations.

In this example, a dataset of 60,000 PD records contained free-text descriptions that require manual classification into 25 subcategories.<sup>26</sup> LLMs, such as Generative Pre-trained Transformer (GPT)-based systems, can analyze this data by leveraging pre-trained language models to understand the context and semantics of the text. These models can classify PDs directly or can generate structured labels that map deviations to their respective subcategories, significantly reducing the reliance on manual review (e.g., free-text description of “participant missed Visit 3 due to transportation issues” could be mapped to a category of “Missed Visit/Visit Out of Window”).

A practical workflow involves using LLMs for initial processing, where the models identify patterns, extract relevant features, and map deviations to predefined subcategories based on their textual content. This can be further enhanced by integrating techniques such as document-term matrices and word embedding models like Word2Vec to preprocess and enrich the dataset. These transformations ensure the data are in a format suitable for traditional ML classifiers. Subsequently, human expertise is applied to validate and refine the training dataset, creating a robust feedback loop. For instance, an LLM could replace or could complement traditional NLP and shallow ML methods, such as support vector machines, which previously achieved an 84% classification accuracy in a similar scenario. The LLM’s ability to contextualize and understand nuanced language would likely improve accuracy and expand applicability, particularly for rare or ambiguous deviations. Additionally, LLMs can provide real-time insights by identifying trends and anomalies in PD data, enabling proactive measures to mitigate risks.

By integrating LLMs into the protocol deviation workflow, clinical trial teams can reduce the time and effort required for manual classification, improve the accuracy and granularity of PD categorizations, and ensure that critical safety and compliance issues are addressed promptly. This not only enhances operational efficiency but also fosters a higher standard of patient safety and data quality in clinical research.

**Case 4: Use of External Control Arms**

Randomized controlled trials are the gold standard for assessing the efficacy of an intervention.<sup>27</sup> However, when randomization is not possible, such as for long-term or rare outcomes, the concept of an external control arm offers an alternative in the study design.<sup>28</sup> Additionally, while randomization may be possible, it can sometimes be impractical or overly costly, particularly when robust data is already available that satisfies the defined criteria, including the standard of care. This involves using individual patient-level data from historical trials in the same indication where subjects meet similar eligibility criteria and baseline demographic and disease characteristics that statistically match those in the experimental arm of the current trial. AI/ML can identify patients who meet eligibility criteria and apply propensity score matching to balance baseline characteristics.<sup>29,30</sup> A challenge with the propensity score approach is in identifying enough historical patients who precisely meet these conditions to allow for statistically valid conclusions. Another issue that arises when trying to use historical patient data is that it does not consider the nuanced changes that occur in patient populations over time, and the various factors embedded in a clinical trial that have a non-trivial effect on drug and placebo response. These factors include the protocol, interactions with the scientific and medical staff running the trial at the various sites, and other uncontrollable conditions (e.g., environmental and political disruptions).

In a case study, the effectiveness of an AI/ML approach was tested using historical non-small cell lung cancer clinical trials.<sup>31</sup> One trial was selected as the target trial and other trials were used to build the external control arm. This methodology was employed to select patients who met the key eligibility criteria and required the appropriate study treatment to address the research question. This yielded overlapping survival curves, insignificant log-rank test results, and hazard ratios approximating one, suggesting similar outcomes between the external control arm to those in the target trial.<sup>31</sup> This case study highlights the potential applicability of AI/ML derived models to efficiently overcome challenges faced in clinical trial analysis when randomization is not possible.

**Case 5: Streamlining the Complaint Handling Process with ML Algorithms**

Complaint handling in the pharmaceutical industry is a resource-intensive process, involving data entry of the complaint, manual review, and the categorization of complaints in free-text systems. Such information is used to determine further action, including if regulatory notification is needed. AI/ML algorithms can help to automate these processes.<sup>32</sup>

In this example, approximately 19,000 monthly complaints spanning two years and encompassing 16,000 products were used as training data to automate the complaint handling process, with the goal of predicting product experience codes and fully integrating this information with the client's enterprise complaint management system. At first, a random forest algorithm was utilized, but later deep learning sequential algorithms

were utilized using the TensorFlow framework in order to overcome memory issues. The resulting model achieved accuracy rates that ranged from 86% to 98% for different product experience codes, with an overall accuracy of 92%. An intuitive user interface was also developed to allow users to easily access and apply the recommendations and predictions.<sup>20</sup> This use case exemplifies how AI/ML can streamline complaint handling, improve accuracy, and reduce resource demands.

**Case 6: AI for Patient Stratification in the Diagnosis Process**

Text analytics, facilitated by AI/ML, have the potential to refine information processing for patient stratification during the diagnostic process in clinical practice and trials. For example, clinicians can copy and paste unstructured data from medical records into commercially available software which quickly and effectively extracts and organizes relevant information according to parameters such as medical, laboratory, and genetic tests and values.<sup>33</sup>

In this case example, an AI/ML system was developed to support and expedite rare disease diagnosis.<sup>34–36</sup> This platform employs neural networks for advanced phenotyping to match symptom descriptions to Human Phenotype Ontology terms, and optical character recognition for text extraction from images. It then uses Exomiser, a Java program that finds potential disease-causing variants from whole-exome or whole-genome sequencing data, to organize and evaluate patient mutation data. The machine then aggregates multiple pathogenicity scores and filters to suggest potential differential diagnoses to the clinician. The result is a stratified view of patients based on likely diagnostic profiles. By leveraging these AI/ML-driven insights, clinicians can identify biologically meaningful subpopulations with distinct diagnostic signatures. This stratification enables more accurate and timely diagnosis and supports targeted treatment pathways for those most likely to respond, increasing the effectiveness of therapeutic interventions and enriching the evidence base in clinical research.<sup>34</sup>

**Case 7: AI-Enhanced Patient Enrichment for Placebo and Drug Response**

Pharmaceutical companies employing control arms in clinical trials face the challenge of placebo response confounding drug response results.<sup>37</sup> Particularly in psychiatric trials, specific clinical scales are rich in psychological and attitudinal insights that enable AI to create placebo response models that can be used to enrich future trials.<sup>38</sup>

In one early example, a set of scales including the Montgomery-Asberg Depression Rating Scale, Beck's Depression Inventory, and Hamilton Anxiety Rating Scale were used in a bipolar depression trial.<sup>38</sup> A classical ML method was used to segment the trial participant population and reveal explainable factors for placebo response. The resulting subpopulations were then used to train a mathematically augmented ensemble tree model to distinguish placebo responders from non-responders. The mathematical augmentation involved a specialized geometric embedding that introduces a kind

of distance between patients. Validation was performed on a completely separate patient trial and demonstrated generalizability with an 87% accuracy. This model was able to categorize trial participants as either placebo responders, placebo non-responders, and unknown.

Recent advances in ML have extended shallow methods, such as ensemble trees, through integration with LLMs and novel mathematical approaches tailored to small data sets.<sup>39</sup> Small data sets pose significant challenges due to their incomplete representation of underlying patient population distributions. However, emerging mathematically augmented techniques enable the identification of subpopulations with high effect sizes, offering critical insights into which patients are most explainable and how they can be prioritized using enrichment criteria.

These advanced methods rank patient subpopulations by their potential to optimize drug response rates, while LLMs augment this process with qualitative insights to refine and enhance the rankings. When applied early, such as in Phase 2 trials, these approaches provide actionable criteria for subsequent trial phases. The derived insights inform the design of inclusion and exclusion criteria, aiming to reduce placebo responses and maximize drug efficacy in follow-on trials. Importantly, the contemporaneous use of small data sets avoids biases introduced by older or external datasets, preserving the trial's relevance and specificity.

By leveraging AI/ML to identify subpopulations with high effect sizes, clinical trialists gain insights into the most pertinent factors driving responses in their study population. These methods allow for a finely tuned balancing act: simultaneously selecting for patients who are unlikely to benefit from the control arm but are predicted to preferentially respond to the active treatment. This dual focus ensures that the trial design maximizes the therapeutic signal while maintaining robust and unbiased results.

Collectively, these seven use cases serve to provide an initial glimpse into how AI/ML could influence clinical trials, underscoring the multifaceted role of AI/ML in enhancing the efficiency, accuracy, and overall effectiveness of clinical trials—from enhancing data quality and patient safety to streamlining complex processes and enriching the understanding of patient responses. AI/ML technologies are rapidly transitioning from tools of operational efficiency to engines of scientific insight within clinical development. Beyond conventional applications—such as forecasting interim analysis timing or benchmarking site performance—emerging methodologies are transforming how we understand and optimize trials themselves.

One such frontier involves the early identification of clinical trial sites exhibiting anomalous data patterns, specifically patterns that diverge from established clinical expectations or violate normative statistical relationships. By leveraging ML to detect deviations from expected symptom interdependencies and latent variable relationships, researchers can pinpoint sites whose data may undermine trial integrity. These insights allow for

pre-randomization interventions—such as targeted audits, exclusion, or re-stratification—thus preserving statistical power and internal validity.

Further innovations in AI and ML are advancing precision and efficiency in clinical research by enabling the identification of hidden patient subgroups, the early detection of subtle clinical changes through continuous monitoring, and the generation of dynamic synthetic control arms using real-world and historical data. Building on these capabilities, privacy-preserving federated learning extends the reach of AI by enabling insights across decentralized datasets, supporting the development of scalable and generalizable models without compromising data privacy.

As AI/ML continues to evolve, its impact on clinical research promises to be even more profound, offering innovative solutions to longstanding challenges.

### Challenges Associated with the Use of AI/ML in Clinical Trials

The use and application of AI/ML in clinical trials, while promising, faces several challenges that have contributed to its relatively slow adoption.<sup>20,40</sup> Three key challenges are generalizability, provenance, and the necessity for effective clinical trialist-AI/ML interaction.

Generalizability pertains to how well AI models can perform beyond their original training, testing, and validation data. Ensuring an AI derived predictive model can generalize and provide relevant recommendations on previously unseen data is crucial. Achieving generalizability faces several challenges related to:

- the availability of large and diverse datasets;
- data preparation and processing tasks, such as annotation, labelling and enrichment, biases elimination, and design choices;
- assumptions made concerning the data's measurement and representation;
- potential biases stemming from factors like the standard of care, represented population, data quality, and healthcare settings;
- the use of Chain-of-Thought prompting to improve the explainability from LLM outputs and to facilitate the understanding of how these systems are evolving;<sup>41</sup>
- the creation of sophisticated algorithms to address the limitations of methods in vogue now.

The provenance of ML algorithms—the decisions, implicit or hidden, that were made in the creation of the model—poses a challenge in the utility and application of AI/ML in clinical trials. Capturing design constraints a priori is essential, with model performance dictated by what is both safe and useful. Performance criteria that satisfy constraints, such as accuracy and error rates, can and should be defined and explicitly captured in advance. Moreover, understanding the provenance of an AI/ML algorithm—its origin, training data, assumptions, and development process—is essential to ensure its reliability and contextual validity. This includes documenting

the source and structure of the data, the preprocessing steps, the rationale behind model choices, and any transformations or augmentations applied. For instance, a model trained on predominantly North American clinical trial data may underperform in global settings due to demographic or procedural mismatches. Similarly, models trained on historical trial data may inadvertently inherit outdated clinical practices or embedded biases. Clear documentation of provenance supports reproducibility, facilitates regulatory review, and enables ethical evaluation of model impact on patient safety and equity.

In practice, cutting edge methods designed to discover enrichment criteria to de-risk clinical trials explicitly track and expose model provenance at each stage of development, from data lineage through to variable selection and subgroup formation. This approach not only enhances interpretability but also allows trial sponsors and regulators to interrogate how specific data characteristics influence model outputs and subgroup definitions, bringing traceability and accountability into AI-driven trial analytics.

Supervised learning models in clinical trials, reliant on physician- and scientist-provided categorization labels for drug responses and diseases, often reinforce preexisting categorizations rooted in current knowledge. This feedback loop not only propagates errors inherent in the labeling process but also limits the discovery of novel subpopulations, especially in heterogeneous diseases like cancer or psychiatric disorders. Traditional ML methods often oversimplify complex spectra of patient responses, leading to missed insights into alternative mechanisms of action. To address this, leveraging unsupervised or semi-supervised approaches, explainable AI, and multi-modal data integration can uncover hidden patient subgroups and refine our understanding of disease. These methods challenge static categories and offer dynamic insights into patient variability, paving the way for a more nuanced and equitable exploration of clinical trial populations. Further, there is an opportunity to utilize novel mathematical methods to address these challenges.

Following this theme, complex ML algorithms can be resource-intensive, demanding substantial data volumes for effective training. To mitigate this, there are research efforts underway exploring methods to extract valuable insights, even from limited datasets. This involves considering innovative mathematical foundations for algorithms and improved collaboration between AI systems and clinical trialists. These challenges are particularly pertinent when explainability of AI systems is limited. While transparency in AI operation is valuable for assisting users in making informed decisions, non-transparent methods can still provide insights, particularly in complex domains like clinical trials. The limitations of these “black-box” approaches can be mitigated by incorporating human-in-the-loop systems, ensuring that expert feedback guides and validates AI outputs. Additionally, complementing these methods with explainable AI enhances interpretability, enabling users to balance the predictive power of opaque models with the actionable insights of more transparent techniques.

## Ethical Considerations for Trustworthy AI in Clinical Trials

The integration of AI into clinical trials presents ethical considerations that must be addressed to ensure the trustworthiness and safety of these technologies for their intended purpose. The evolving field of ethical AI underscores the need to protect patients' rights, privacy, and safety.<sup>40</sup> Establishing a robust and trustworthy foundation is therefore crucial for the integration of AI in clinical trials.

While de-identification is a standard practice in clinical research, recent advances in AI have raised legitimate concerns about the potential for re-identifying individuals within ostensibly anonymized datasets. Complex AI models, particularly those trained on large and diverse data sources, can sometimes detect subtle patterns that correlate across datasets, unintentionally increasing the risk of re-identification. This possibility necessitates the implementation of robust, multi-layered privacy controls, including techniques such as differential privacy, federated learning, and rigorous access governance.<sup>42,43</sup> Addressing these risks is critical for maintaining patient trust and upholding ethical standards in pharmaceutical research. The primary objective for developing AI-based applications for clinical trials is to create products with trustworthy design, development, and testing processes. This is vital for societal trust in the AI-based product by all interested parties, including patients, clinical trialists, and regulators. AI/ML approaches have shown promise in proof-of-concept and academic studies and have recently found themselves being used in actual clinical trials.<sup>44</sup>

A risk involved is that AI systems may produce models that are driven by erroneous factors. For example, an AI system designed to diagnose cancer lesions performed well in a proof-of-concept study but failed in a real-world setting. In this instance, the AI mistakenly learned to rely on human-derived measurements, such as rulers placed next to tumors in images, as the primary indicator to classify cancer images.<sup>45,46</sup> The resulting model erroneously used these adjacent images as cancer identifiers rather than assessing the lesion itself.<sup>47</sup> This example illustrates that AI/ML needs to be trained according to a rigorous set of standards, and that for medical purposes, even though explainability may not be strictly required, as described previously in this paper, effective human-AI collaborative practices need to be employed.<sup>48</sup>

Despite recent advances, the EU Commission and the High-Level Expert Group on Artificial Intelligence (AI HLEG) have prudently published guidelines for Ethical and Trustworthy AI.<sup>49</sup> These guidelines emphasize a human-centric approach and outline seven core requirements that AI systems should meet to be considered trustworthy. The guideline states that throughout the AI system's entire life cycle, trustworthy AI should be lawful, ethical, and robust (from both technical and social perspectives). Within the guidelines, four ethical principles are considered ethical imperatives within the context of AI: respect for human autonomy, prevention of harm, fairness, and explicability. Overall, the outputs by the AI

HLEG have served as resources to multiple policy-making initiatives in this area.<sup>49</sup>

The EU has introduced the 'Artificial Intelligence Act' as a set of regulations to ensure the ethical use of AI.<sup>50,51</sup> This legislation aims to promote the development of trustworthy AI systems with a focus on protecting the fundamental rights of citizens, building public trust in AI, and promoting its widespread adoption. The Artificial Intelligence Act adopts a risk-based approach, classifying AI systems as unacceptable, high-risk, limited-risk, or minimal-risk. It outlines obligations for these systems including using adequate risk and quality management systems; providing clear, concise, and transparent instructions for use; maintaining high-quality datasets for training, validation, and testing; and the allowance for human override capabilities.

The ethical considerations surrounding AI in clinical trials are crucial in ensuring safety, trustworthiness, and compliance with regulations. Adhering to ethical principles and guidelines is essential to harness the potential of AI while safeguarding patients and upholding societal values. It is also hoped that by providing legal certainty on the permitted use of AI, these guidelines will encourage innovation and investment in the sector.

### Good Machine Learning Practices in Clinical Development

The adoption of Good ML Practices (GMLP) is essential to ensure the effective and ethical use of AI/ML, particularly in clinical development use cases. The adoption of GMLP is influenced by several factors, such as transparency and explainability with proper documentation; data quality and relevance; performance monitoring; and validation. These offer a common-sense framework applicable to scenarios where the ML methods are complex and it is difficult to directly assess the logic of their outputs, such as those involving large neural networks.<sup>52</sup>

Defining the intended scope of the use case is paramount and requires the provision of comprehensive and contextually relevant information regarding the ML model's performance, encompassing details about the training and testing data, acceptable inputs, known limitations, how to interpret results, and model integration into the overall solution. Additionally, feature engineering—the process of selecting, manipulating, and transforming raw input data into features that can be used in ML—plays a vital role. GMLP places significant emphasis on the quantity and quality of training data, as model accuracy heavily relies on input features, feature importance, and diversity of training data.<sup>53</sup> Best practices in GMLP should ensure independence of training and test datasets to mitigate bias and confounding factors. Feature importance assigns higher significance to some specific input features, which enhances model generalization across larger datasets (e.g., higher feature importance of females in breast cancer prediction models).<sup>54</sup>

When designing a model, it is crucial to consider its intended use and to mitigate risks of overfitting to training data. Utilizing human-in-the-loop methodologies provide

useful checks and balances on machine predictions, improving the model's learning capabilities through user feedback. Deep learning models have reduced the need for manual feature engineering by leveraging representation learning to extract relevant features directly from raw data. However, traditional machine learning models, such as decision trees or logistic regression, still depend heavily on well-defined input features, making feature engineering crucial for optimizing their performance. It is important to distinguish between these approaches, as preprocessing and light feature engineering can still enhance the performance of deep learning models, particularly when applied to structured data, where domain knowledge can guide the model toward better representations.

It is also important to note that in clinical trials, particularly during enrichment, it is essential for AI to provide a clear and actionable prescription of the features and their specific ranges that define a superior pre-randomization cohort. This level of precision ensures that the identified patient groups are optimized for achieving meaningful trial outcomes. For such use cases, the use of simulations becomes critical, enabling robust evaluation of potential cohorts under varying scenarios. Equally important is the AI's ability to identify subpopulations with sufficient effect sizes, allowing for the recommendation of effective cohorts even when working with the inherently small datasets typical of clinical trials.

Collectively, GMLP serve as a comprehensive guide to ensure the effective and ethical application of AI/ML in clinical development.<sup>52</sup> They emphasize the importance of defining a clear use case, providing transparency in model performance, and competent feature engineering. The quality and independence of training and testing datasets are paramount to the model's accuracy and generalizability. Furthermore, the integration of human-in-the-loop methodologies not only provides a safety net for machine predictions but also enhances the model's learning capabilities through a feedback mechanism. GMLP also underscore the necessity of balancing model complexity with interpretability, ensuring transparency in predictions while harnessing the full potential of AI/ML. Adhering to these practices is essential for mitigating risks, ensuring fairness, and maintaining trust in advanced AI technologies.<sup>52</sup> **Table 2** summarizes key GMLP considerations for the implementation of AI/ML in clinical trials.

### Regulatory Considerations for the Use of AI/ML in Clinical Trials

Regulators recognize the potential of AI/ML technologies in advancing drug development and streamlining clinical trials and have taken proactive steps, formulating strategies, action plans, and informational documents that provide guidance on the use of AI/ML-based software in medical devices and drug development.<sup>55</sup> Recognizing the need for responsible use, regulatory agencies actively engage in outreach to develop principles and guidance addressing the unique challenges and risks associated

**Table 2:** GMLP considerations for AI/ML implementation in clinical trials.

<b>GMLP Considerations</b>	<b>Description</b>
<b>Document Process</b>	Thoroughly document training, validation, and testing phases, including deviations from the plan.
<b>Ensure Data Integrity</b>	Use large, diverse datasets with proper identification and storage for training, validation, and testing.
<b>Avoid Data Overlap</b>	Verify test data independence, ensuring it doesn't overlap with training or validation data.
<b>Rigorous Test Data Selection</b>	Separate test data carefully and select it based on relevant criteria for representativeness and challenge.
<b>Robust Data Processing</b>	Apply appropriate cleaning, normalization, and exclusion criteria to maintain data quality in test data.
<b>Feature Analysis</b>	Understand the impact of features on the algorithm's output and select relevant test data accordingly.
<b>Account for Technical Differences</b>	Ensure test data covers potential real-world variations in formatting and data sources.
<b>Verify Classifications</b>	Validate correctness of data classifications, possibly involving second-person verification or lab tests.
<b>Keep Data Up-to-date</b>	Regularly assess test data relevance and plan for retraining to address data changes.
<b>Address Bias and Variance</b>	Optimize the algorithm to balance bias and variance tradeoff and assess results using graphs.
<b>Use Appropriate Metrics</b>	Utilize metrics like Sensitivity, Specificity, Precision, and F1 Score for evaluating model performance.
<b>Focus on Key Metrics</b>	Emphasize relevant metrics and confusion matrix quadrants based on the application's scope.
<b>Define Application Scope</b>	Limit the scope based on test data and results, ensuring the algorithm's applicability.
<b>Set Appropriate Thresholds</b>	Establish clear thresholds for end results and determine when human interaction is required for certain outcomes.

with the responsible use of AI/ML in clinical trials.<sup>8,56</sup> An example of such outreach efforts is the EMA's GCP Inspector Work Group stakeholder meetings held in 2020<sup>20</sup> and 2021<sup>40,57</sup>.

Moreover, regulatory agencies strongly encourage sponsors and other interested parties to initiate early and frequent communication, particularly when employing AI/ML in clinical development, such as for study population enrichment, assessment of endpoints, and to inform study design. The 2021 stakeholder meeting discussed avenues for regulatory engagement, soliciting input and advice, as well as criteria used to evaluate AI/ML technology. Currently, there are several established forums for engagement with regulatory agencies,<sup>8,58-61</sup> as well as published guidance documents, action plans, and other informational documents regarding the integration of new technologies, including AI/ML in clinical trials. However, summarizing the details provided in these documents is beyond the scope of this paper.<sup>52,53,62-76</sup>

Existing regulatory frameworks lack specific requirements or guidance documents pertaining to the evaluation of AI/ML algorithms in critical GCP applications. However, the Danish Medicines Agency (DKMA) published a proposal in March 2021 that includes pointed questions aimed at comprehensively assessing AI/ML model risks and accuracy.<sup>77</sup> These questions probe the design, training, validation, and testing processes, with a focus on identifying potential issues in machine training data, including biases, validation challenges, and

algorithm accuracy. The DKMA proposal scrutinizes model selection and optimization and evaluates the relevance, sufficiency, and integrity of test data and associated outcomes.<sup>77</sup> It is important to note that the DKMA proposal, while informative, may not necessarily reflect the official position of the EMA or the GCP IWG.

In general, regulators recommend a transparent, risk-proportionate approach to the management of AI/ML technologies throughout the entire clinical trial lifecycle.<sup>55</sup> This approach encompasses comprehensive documentation of the algorithm and its intended use, adherence to GMLP during model creation and validation, and the careful evaluation of training, testing, and validation datasets. In addition, it is imperative to maintain and retain necessary documentation throughout the AI/ML technologies' lifecycle.

### **Future Directions of the Use of AI/ML in Clinical Trials**

With respect to clinical trials for new drug development, the integration of AI/ML applications with privacy-preserving features is emerging as a promising avenue. One notable approach is federated or collaborative learning, which enables the training of data models locally without the need for actual data sharing.<sup>78</sup> This approach proves advantageous by allowing multiple organizations to collaboratively train AI/ML models on their respective datasets while safeguarding sensitive patient information and proprietary data. Instead of sharing raw data,

federated learning aggregates learned model updates, enhancing the overall performance and generalization of AI/ML models. This approach not only addresses privacy concerns but also allows researchers to leverage larger and more diverse datasets, ultimately leading to more robust and precise AI-driven insights in drug development. Such collaborative endeavours hold the potential to expedite the discovery of novel therapies, optimize clinical trial designs, and improve patient outcomes.

Furthermore, the advancement of LLMs offers additional avenues for enhancing and streamlining clinical trials.<sup>79</sup> For example, LLMs, and their fusions with sophisticated ML and heuristics, can analyze EHRs to identify potential trial participants who meet specific study eligibility criteria to facilitate the use of RWD and patient recruitment in clinical trials.<sup>80,81</sup> These models may streamline the collection of RWD in clinical trials by automatically extracting relevant data—such as standard of care and medical history from the EHRs—and inputting it directly into EDC systems.<sup>82</sup> They may also support the design of future trials by identifying patterns such as drug response profiles, placebo sensitivity, adverse event risk factors, and potential drug interactions, based on insights learned from previously completed studies. These insights can inform the development of more effective enrichment strategies, including refined inclusion and exclusion criteria, all intended for use in the pre-randomization phase to enhance trial efficiency and increase the likelihood of detecting true therapeutic effects.<sup>81,83</sup> Additionally, LLMs may assist in developing personalized messaging strategies to keep patients engaged in the trial and motivated to complete them successfully. Finally, by having LLMs interpret results from emerging sophisticated explanatory ML algorithms that learn from clinical trial data, one may benefit from having insights interpreted through a large corpus of medical literature. These hybrid systems can inform future trial design, such as through identifying patient subpopulations with differential drug responses or by identifying confounding factors that may need to be controlled for in future trials. These advancements not only promise increased efficiency in clinical trial processes but also hold the potential to improve the drug development landscape by accelerating discoveries and improving patient outcomes.

Another promising direction involves constructing a federation of algorithms capable of supervised learning, augmented by human expertise and unsupervised methods. This approach enables clinical trialists to optimize their trials through personalization and adverse event modelling, identifying patient subpopulations best served by specific treatments. Further, emerging developments suggest that Agentic AI coupled with sophisticated mathematical augmentation, holds the promise of reshaping clinical trials by dynamically adapting to trial complexities, identifying actionable subpopulations in real-time, and offering predictive insights that accelerate drug development while ensuring patient safety and efficacy.

The accelerating impact of AI/ML in clinical research and medicine is driven by several converging mechanisms. First, the growing availability of high-dimensional

real-world and clinical trial data, spanning genomics, imaging, behavioral metrics, and EHRs, enables richer and more representative model development. Second, advances in model architectures, particularly those tailored for small, heterogeneous datasets, now allow for interpretable subpopulation discovery and hypothesis generation in early-phase and rare disease trials.<sup>39,84</sup> These efforts are leading to frameworks that retain full traceability from source data to insight, ensuring clinical relevance while maintaining auditability. Finally, government organizations are advancing frameworks for assessing AI/ML model risk, robustness, and bias, laying the groundwork for safe and scalable deployment.<sup>50,85</sup> These developments, taken together, make a compelling case that AI/ML will not merely support, but will actively shape, the future of clinical trial design and therapeutic decision-making.

## Conclusion

As AI/ML continues to reshape clinical trials, industry and regulatory authorities are rapidly adapting to these changes. Stakeholder meetings play a crucial role in fostering communication, discussing challenges, and seeking scientific advice. The potential of AI/ML offers both opportunities and challenges, pushing the boundaries of innovation to explore novel solutions. Success hinges on building trust through open communication, transparency, and realistic expectations with regulators and other interested parties, including patient advocacy groups.

For regulators and clinical trial leadership to fully embrace AI/ML advancements, transparency, along with the safe and effective development and implementation of these technologies, is key. This includes safeguarding participants' rights and safety, ensuring data quality and integrity, and driving efficiency improvements that go beyond merely digitizing existing processes. AI/ML opens doors to new approaches, such as federated learning, enabling insights from external data without centralization. The unique requirements of clinical trials are pushing AI/ML to innovate further so that systems can learn from smaller data and that explainability becomes a priority.

## Disclaimer

This article reflects the views of the authors and may not be understood or quoted as being made on behalf or reflecting the position of the agencies or organizations with which the authors are affiliated.

## Acknowledgements

The authors thank the EMA Good Clinical Practice Inspectors Working Group for their efforts in hosting the two virtual conferences in 2020 and 2021. Thanks also go to the Society for Clinical Data Management (SCDM) for their logistical support in the initial summary of the first conference discussions. The following individuals are acknowledged for their contribution in the organization, presentation and/or discussion in the virtual conference session(s): Camelia Mihaescu (EMA), Jane Moseley (EMA), Ashley Howard (Pfizer), Willie Muehlhausen (Safria Clinical

Research), Melissa Binz (Pfizer), Emma Richard (Johnson & Johnson), Ruthie Davi (Acorn AI), Julián Isla (Foundation 29), Kevin Lyman (Enlytic), Bruno Boulanger (PharmaLex), Matthew Diamond (US FDA), Ivan Walrath (Pfizer), Robert Vandersluis (GSK), Fiona Maini (Medidata Solutions), Xiaoxuan Liu (University Hospitals Birmingham), Mihaela Van Der Schaar (University of Cambridge), Kim Branson (GSK), Ilan Halberstam (Idorsia), Jesper Kjaer (DKMA), Ib Alstrup (DKMA), Steven Berman (US FDA), Dennis Bergau (Abbvie), Yiannos Toliás (European Commission) and Jelena Malinina (European Consumer Organization).

### Competing Interests

Drs. Geraci and Qorri are employees of NetraMark Corp. Dr. Geraci is a significant shareholder of NetraMark Corp. which develops commercial clinical trial optimization and precision medicine products. Dr. Geraci is also affiliated with the Department of Molecular Medicine and Pathology, Queen's University, Kingston, Ontario, Canada; Center for Biotechnology and Genomic Medicine, Augusta University, Georgia, USA; and the Centre for Addiction and Mental Health, Toronto, Canada; Arthur C. Clarke Center for Human Imagination, School of Physical Sciences, University of California, San Diego, CA, USA. Mr. Rao is an employee of Saama Technologies, and he participated in the EMA-GCP IWG AI in clinical trials workshops when he was employed by Pfizer. Mr. Nadolny is a full-time employee of Sanofi and a member representative of the Society for Clinical Data Management (SCDM). Dr. Edwards is an employee of Relation. Ms. Hofmann is an employee of Cognizant Technology Solutions. Dr. Khozin participated the EMA-GCP IWG AI in clinical trials workshops when he was employed by Johnson & Johnson, Inc. and ASCO's CancerLinQ LLC., in 2020 and 2021, respectively. Currently, he is a Research Affiliate at the MIT, and Principal, PhusionBio, LLC. Mr. Schaltenbrand is an employee of Wega Informatik AG. Mr. Yeomans is an employee of Viedoc Technologies. Mr. Zambas is an employee of Pfizer, Inc. NYC, USA and is a member representative of SCDM. Dr. Khin participated in the EMA GCP-IWG working group AI in clinical trials workshop planning activities when she was previously employed by the US Food and Drug Administration. Currently, she is employed by Neurocrine Biosciences, Inc.

### References

1. **Bohr A, Memarzadeh K.** The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*. Published online January 1, 2020;25–60. DOI: <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>
2. **Bajwa J, Munir U, Nori A, Williams B.** Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J.* 2021;8(2):e188–e194. DOI: <https://doi.org/10.7861/fhj.2021-0095>
3. **Johnson KB, Wei WQ, Weeraratne D,** et al. Precision medicine, AI, and the future of personalized health care. *Clin Transl Sci.* 2021;14(1):86–93. DOI: <https://doi.org/10.1111/cts.12884>
4. **Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK.** Artificial intelligence in drug discovery and development. *Drug Discov Today.* 2021;26(1):80–93. DOI: <https://doi.org/10.1016/j.drudis.2020.10.010>
5. **Askin S, Burkhalter D, Calado G, El Dakrouni S.** Artificial intelligence applied to clinical trials: opportunities and challenges. *Health Technol (Berl).* 2023;13(2):203–213. DOI: <https://doi.org/10.1007/s12553-023-00738-2>
6. **European Medicines Agency.** Annual Report of the Good Clinical Practice Inspectors Working Group 2020. Published online 2020. Accessed September 24, 2023. [https://www.ema.europa.eu/en/documents/report/annual-report-good-clinical-practice-inspectors-working-group-2020\\_en.pdf](https://www.ema.europa.eu/en/documents/report/annual-report-good-clinical-practice-inspectors-working-group-2020_en.pdf)
7. **European Medicines Agency.** Annual Report of the Good Clinical Practice Inspectors Working Group 2021. Published online 2021. Accessed January 7, 2025. [https://www.ema.europa.eu/en/documents/report/annual-report-good-clinical-practice-inspectors-working-group-2021\\_en.pdf](https://www.ema.europa.eu/en/documents/report/annual-report-good-clinical-practice-inspectors-working-group-2021_en.pdf)
8. **Liu Q, Huang R, Hsieh J,** et al. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. *Clin Pharmacol Ther.* 2023;113(4):771–774. DOI: <https://doi.org/10.1002/cpt.2668>
9. **Vaswani A, Shazeer NM, Parmar N,** et al. Attention is all you need. *Neural Information Processing Systems*. Published online 2017. DOI: <https://doi.org/10.48550/arXiv.1706.03762>
10. **Quantiphi.** From Data to Drugs: The Promising Intersection of Generative AI and Pharma Industry. Published July 7, 2023. Accessed May 21, 2025. <https://quantiphi.com/from-data-to-drugs-the-promising-intersection-of-generative-ai-and-pharma-industry/>
11. **Sarker IH.** Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci.* 2021;2(3):160. DOI: <https://doi.org/10.1007/s42979-021-00592-x>
12. **Liu F, Demosthenes P.** Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol.* 2022;22(1):287. DOI: <https://doi.org/10.1186/s12874-022-01768-6>
13. **Weissler EH, Naumann T, Andersson T,** et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials.* 2021;22(1):537. DOI: <https://doi.org/10.1186/s13063-021-05489-x>
14. **Madan S, Lentzen M, Brandt J, Rueckert D, Hofmann-Apitius M, Fröhlich H.** Transformer models in biomedicine. *BMC Medical Informatics and Decision Making.* 2024;24(1):1–22. DOI: <https://doi.org/10.1186/s12911-024-02600-5>
15. **Denecke K, May R, Rivera-Romero O.** Transformer models in healthcare: a survey and thematic analysis of potentials, shortcomings and risks. *J Med Syst.* 2024;48(1):23 DOI: <https://doi.org/10.1007/s10916-024-02043-5>

16. **Kumar Y, Koul A, Singla R, Ijaz MF.** Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput.* 2023;14(7):8459–8486. DOI: <https://doi.org/10.1007/s12652-021-03612-z>
17. **Khazin S.** From organs to algorithms: Redefining cancer classification in the age of artificial intelligence. *Clin Transl Sci.* 2024;17(9):e70001. DOI: <https://doi.org/10.1111/cts.70001>
18. **Raju GK, Khazin S, Gurumurthi K, Domike R, Woodcock J.** Patient-centered approach to benefit–risk characterization using number needed to benefit and number needed to harm: advanced non–small-cell lung cancer. *JCO Clin Cancer Inform.* 2020;4(4):769–783. DOI: <https://doi.org/10.1200/CCI.19.00103>
19. **U.S. Food and Drug Administration.** FDA approves first cancer treatment for any solid tumor with a specific genetic feature. Published May 23, 2017. Accessed September 24, 2023. <https://www.fda.gov/news-events/press-announcements/fda-approves-first-cancer-treatment-any-solid-tumor-specific-genetic-feature>
20. **European Medicines Agency.** Artificial intelligence in clinical trials – ensuring it is fit for purpose. YouTube. Accessed August 13, 2024. <https://www.youtube.com/watch?v=T92f809QIGU>
21. **Pfizer.** How a Novel ‘Incubation Sandbox’ Helped Speed Up Data Analysis in Pfizer’s COVID-19 Vaccine Trial. Accessed September 25, 2023. [https://www.pfizer.com/news/articles/how\\_a\\_novel\\_incubation\\_sandbox\\_helped\\_speed\\_up\\_data\\_analysis\\_in\\_pfizer\\_s\\_covid\\_19\\_vaccine\\_trial](https://www.pfizer.com/news/articles/how_a_novel_incubation_sandbox_helped_speed_up_data_analysis_in_pfizer_s_covid_19_vaccine_trial)
22. **Izmailova ES, Wagner JA, Perakslis ED.** Wearable devices in clinical trials: hype and hypothesis. *Clin Pharmacol Ther.* 2018;104(1):42–52. DOI: <https://doi.org/10.1002/cpt.966>
23. **Mitsi G, Grinnell T, Giordano S, et al.** Implementing digital technologies in clinical trials: lessons learned. *Innov Clin Neurosci.* 2022;19(4–6):65–69.
24. **Coravos A, Khazin S, Mandl KD.** Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digital Medicine* 2019 2:1. 2019;2(1):1–5. DOI: <https://doi.org/10.1038/s41746-019-0090-4>
25. **Brophy E, Muehlhausen W, Smeaton AF, Ward TE.** Optimised convolutional neural networks for heart rate estimation and human activity recognition in wrist worn sensing applications. Published online March 30, 2020. Accessed September 27, 2023. <https://arxiv.org/abs/2004.00505v1>
26. **Richard E, Reddy B.** Text classification for clinical trial operations: evaluation and comparison of natural language processing techniques. *Ther Innov Regul Sci.* 2021;55(2):447–453. DOI: <https://doi.org/10.1007/s43441-020-00236-x>
27. **Akobeng AK.** Understanding randomised controlled trials. *Arch Dis Child.* 2005;90(8):840–844. DOI: <https://doi.org/10.1136/adc.2004.058222>
28. **Thorlund K, Dron L, Park JJH, Mills EJ.** Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol.* 2020;12:457–467. DOI: <https://doi.org/10.2147/CLEP.S242097>
29. **Austin PC.** An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* 2011;46(3):399–424. DOI: <https://doi.org/10.1080/00273171.2011.568786>
30. **Corder N, Yang S.** Utilizing stratified generalized propensity score matching to approximate blocked randomized designs with multiple treatment levels. *J Biopharm Stat.* 2022;32(3):373–399. DOI: <https://doi.org/10.1080/10543406.2022.2065507>
31. **Yin X, Mishra-Kalyan PS, Sridhara R, Stewart MD, Stuart EA, Davi RC.** Exploring the potential of external control arms created from patient level data: A case study in non-small cell lung cancer. *J Biopharm Stat.* 2022;32(1):204–218. DOI: <https://doi.org/10.1080/10543406.2021.2011901>
32. **Society for Clinical Data Management.** Introduction to Artificial Intelligence in Drug Development (Part 1 and 2). Accessed December 16, 2023. <https://learning-scdm.org/courses/28689>
33. **Huang J, An A, Hu V, Tu K.** Medical text analytics tools for search and classification. *Stud Health Technol Inform.* 2009;143:519–524. DOI: <https://doi.org/10.3233/978-1-58603-979-0-519>
34. **Zhao M, Havrilla JM, Fang L, et al.** Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genom Bioinform.* 2020;2(2). DOI: <https://doi.org/10.1093/nargab/lqaa032>
35. **Foundation 29.** Accessed December 16, 2023. <https://foundation29.org/#home>
36. **Dx29.** Accessed December 16, 2023. <https://dx29.ai/>
37. **Hall KT, Loscalzo J.** Drug-placebo additivity in randomized clinical trials. *Clin Pharmacol Ther.* 2019;106(6):1191–1197. DOI: <https://doi.org/10.1002/cpt.1626>
38. **Smith EA, Horan WP, Demolle D, et al.** Using artificial intelligence-based methods to address the placebo response in clinical trials. *Innov Clin Neurosci.* 2022;19(1–3):60–70.
39. **Geraci J, Bhargava R, Qorri B, et al.** Machine learning hypothesis-generation for patient stratification and target discovery in rare disease: our experience with Open Science in ALS. *Front Comput Neurosci.* 2023;17:1199736. DOI: <https://doi.org/10.3389/fncom.2023.1199736>
40. **GCP IWG.** 2021 Virtual Workshop of the GCP IWG on Artificial Intelligence in Clinical Trials day1 on Vimeo. Accessed August 13, 2024. <https://vimeo.com/video/626397529>
41. **Wei J, Wang X, Schuurmans D, et al.** Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst.* 2022;35. DOI: <https://doi.org/10.48550/arXiv.2201.11903>

42. **Dwork C, Roth A.** The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*. 2014;9(3–4):211–407. DOI: <https://doi.org/10.1561/04000000042>
43. **Brendan McMahan H, Moore E, Ramage D, Hampson S, Agüera y Arcas B.** Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*. Published online February 17, 2016. Accessed April 22, 2025. <https://arxiv.org/abs/1602.05629v4>
44. **Chopra H, Annu, Shin DK, et al.** Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. *Int J Surg*. 2023;109(12):4211–4220. DOI: <https://doi.org/10.1097/JS9.0000000000000705>
45. **Liopyris K, Gregoriou S, Dias J, Stratigos AJ.** Artificial intelligence in dermatology: challenges and perspectives. *Dermatol Ther (Heidelb)*. 2022;12(12):2637–2651. DOI: <https://doi.org/10.1007/s13555-022-00833-8>
46. **Winkler JK, Fink C, Toberer F, et al.** Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatol*. 2019;155(10):1135–1141. DOI: <https://doi.org/10.1001/jamadermatol.2019.1735>
47. **Patel RH, Foltz EA, Witkowski A, Ludzik J.** Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: a systematic review. *Cancers (Basel)*. 2023;15(19):4694. DOI: <https://doi.org/10.3390/cancers15194694>
48. **Daneshjou R, Barata C, Betz-Stablein B, et al.** CheckList for Evaluation of image-based AI Reports in dermatology: CLEAR Derm Consensus Guidelines from the International Skin Imaging Collaboration Artificial Intelligence Working Group. *JAMA Dermatol*. 2022;158(1):90–96. DOI: <https://doi.org/10.1001/jamadermatol.2021.4915>
49. **European AI Alliance Input for the First Workshop of the AI HLEG.** Accessed October 8, 2023. <https://futurium.ec.europa.eu/en/european-ai-alliance/document/european-ai-alliance-input-first-workshop-ai-hleg?language=sk>
50. **The European Union.** Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. Published December 9, 2023. Accessed January 2, 2024. <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>
51. **The European Union.** The Artificial Intelligence Act. Accessed May 27, 2025. <https://artificialintelligenceact.eu/>
52. **U.S. Food and Drug Administration.** Good Machine Learning Practice for Medical Device Development: Guiding Principles. Accessed January 3, 2024. <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles>
53. **U.S. Food and Drug Administration.** Guidance for Industry. Software as a Medical Device (SaMD) Action Plan. Published online January 2021. Accessed January 2, 2024. <https://www.fda.gov/media/145022/download?attachment>
54. **Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L.** Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med Inform Decis Mak*. 2023;23(1):1–14. DOI: <https://doi.org/10.1186/s12911-023-02377-z>
55. **U.S. Food and Drug Administration.** Using Artificial Intelligence & Machine Learning in the Development of Drug and Biological Products. Accessed December 16, 2023. <https://www.fda.gov/media/167973/download>
56. **El Zarrad M, Lee A, Purcell R, Steele S.** Advancing an agile regulatory ecosystem to respond to the rapid development of innovative technologies. *Clinical Translational Science*. 2022;(15):1332–1339. DOI: <https://doi.org/10.1111/cts.13267>
57. **GCP IWG.** 2021 Virtual Workshop of the GCP IWG on Artificial Intelligence in Clinical Trials day2 on Vimeo. Accessed August 13, 2024. <https://vimeo.com/video/629993605>
58. **U.S. Food and Drug Administration.** Critical Path Innovation Meetings (CPIM). 2015. Accessed January 3, 2024. <https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/critical-path-innovation-meetings-cpim>
59. **U.S. Food and Drug Administration.** Critical Path Innovation Meetings Guidance for Industry. Published online 2015. Accessed January 3, 2024. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/formal-meetings-between-fda-and-sponsors-or-applicants-pdufa-products>
60. **U.S. Food and Drug Administration.** Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products. Published September 2023. Accessed January 3, 2024. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/formal-meetings-between-fda-and-sponsors-or-applicants-pdufa-products>
61. **European Medicines Agency.** Reflection paper on the use of Artificial Intelligence (AI). Published online 2023. Accessed January 3, 2024. [https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf)
62. **U.S. Food and Drug Administration.** Biomarker Qualification Program. Accessed January 3, 2024. <https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program>
63. **U.S. Food and Drug Administration.** Qualification Process for Drug Development Tools Guidance

- for Industry and FDA Staff. Published November 2020. Accessed January 3, 2024. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-process-drug-development-tools-guidance-industry-and-fda-staff>
64. **U.S. Food and Drug Administration.** Drug Development Tool Qualification Process: Transparency Provisions. Accessed January 3, 2024. <https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/drug-development-tool-qualification-process-transparency-provisions>
  65. **U.S. Food and Drug Administration.** Innovative Science and Technology Approaches for New Drugs (ISTAND) Pilot Program. Accessed January 3, 2024. <https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program>
  66. **U.S. Food and Drug Administration.** Innovative Science and Technology Approaches for New Drugs (ISTAND) Pilot Program Submission Process. Accessed January 3, 2024. <https://www.fda.gov/drugs/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program-submission-process>
  67. **U.S. Food and Drug Administration.** IStand Qualification Letter of Intent (LOI) Model Content Elements. Accessed January 3, 2024. <https://www.fda.gov/media/142478/download>
  68. **U.S. Food and Drug Administration.** Digital Health Technologies for Remote Data Acquisition in Clinical Investigations. Published December 2023. Accessed January 3, 2024. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations>
  69. **U.S. Food and Drug Administration.** Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)-Discussion Paper and Request for Feedback. Accessed January 3, 2024. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf>
  70. **U.S. Food and Drug Administration.** FDA Releases Artificial Intelligence/Machine Learning Action Plan. Published January 12, 2021. Accessed January 3, 2024. <https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan>
  71. **U.S. Food and Drug Administration.** Digital Health Software Precertification (Pre-Cert) Pilot Program. Published September 26, 2022. Accessed January 3, 2024. <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program>
  72. **U.S. Food and Drug Administration.** Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities: 2020 Update. Accessed May 27, 2025. <https://www.fda.gov/media/142107/download>
  73. **U.S. Food and Drug Administration.** Digital Health Innovation Action Plan. Accessed January 3, 2024. <https://www.fda.gov/media/106331/download>
  74. **IMDRF Software as a Medical Device (SaMD) Working Group.** Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations. Published online 2014. Accessed January 3, 2024. <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf>
  75. **U.S. Food and Drug Administration.** Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions. Accessed May 27, 2025. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial>
  76. **U.S. Food and Drug Administration.** Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products. Accessed April 22, 2025. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological>
  77. **Danish Medicines Agency.** Suggested criteria for using AI/ML algorithms in GxP. Published March 8, 2021. Accessed May 27, 2025. <https://laegemiddelstyrelsen.dk/en/licensing/supervision-and-inspection/inspection-of-authorised-pharmaceutical-companies/using-aiml-algorithms-in-gxp/>
  78. **Guendouzi BS, Ouchani S, EL Assaad H, EL Zaher M.** A systematic review of federated learning: Challenges, aggregation methods, and development tools. *Journal of Network and Computer Applications.* 2023;220:103714. DOI: <https://doi.org/10.1016/j.jnca.2023.103714>
  79. **Clusmann J, Kolbinger FR, Muti HS, et al.** The future landscape of large language models in medicine. *Communications Medicine.* 2023;3(1):141. DOI: <https://doi.org/10.1038/s43856-023-00370-1>
  80. **Park J, Fang Y, Ta C, et al.** Criteria2query 3.0: Leveraging generative large language models for clinical trial eligibility query generation. DOI: <https://doi.org/10.2139/ssrn.4637800>
  81. **Nievas M, Basu A, Wang Y, Singh H.** Distilling large language models for matching patients to clinical trials. *J Am Med Inform Assoc.* 2024;31(9):1953–1963. DOI: <https://doi.org/10.1093/jamia/ocae073>

82. **Datta S, Lee K, Paek H**, et al. AutoCriteria: a generalizable clinical trial eligibility criteria extraction system powered by large language models. *J Am Med Inform Assoc.* 2024;31(2):375–385. DOI: <https://doi.org/10.1093/jamia/ocad218>
83. **Beattie J, Neufeld S, Yang D**, et al. Utilizing Large Language Models for Enhanced Clinical Trial Matching: A Study on Automation in Patient Screening. *Cureus.* 2024;16(5):e60044. DOI: <https://doi.org/10.7759/cureus.60044>
84. **Moses C, Qorri B, Amruth B**, et al. Small Patient Datasets Reveal Genetic Drivers of Non-Small Cell Lung Cancer Subtypes Using Machine Learning for Hypothesis Generation. *Explor Med.* Published online October 2023. DOI: <https://doi.org/10.37349/emed.2023.00153>
85. **Tabassi E.** Artificial Intelligence Risk Management Framework (AI RMF 1.0). Published online January 26, 2023. DOI: <https://doi.org/10.6028/NIST.AI.100-1>

**How to cite this article:** Geraci J, Rao P, Grandinetti C, Qorri B, Nadolny P, Ayalew K, Bregnhøj L, Edwards L, Hofmann K, Khozin S, Schaltenbrand N, Stemmler T, Yeomans A, Zambas D, Khin NA. Current Opportunities for the Integration and Use of AI/ML in Clinical Trials: Good Clinical Practice Perspectives. *Journal of the Society for Clinical Data Management.* 2025; 5(1): 8, pp.1–15. DOI: <https://doi.org/10.47912/jscdm.426>

**Submitted:** 06 March 2025

**Accepted:** 16 May 2025

**Published:** 05 June 2025

**Copyright:** © 2025 SCDM publishes JSCDM content in an open access manner under a Attribution-Non-Commercial-ShareAlike (CC BY-NC-SA) license. This license lets others remix, adapt, and build upon the work non-commercially, as long as they credit SCDM and the author and license their new creations under the identical terms. See <https://creativecommons.org/licenses/by-nc-sa/4.0/>.



*Journal of the Society for Clinical Data Management* is a peer-reviewed open access journal published by Society for Clinical Data Management.

**OPEN ACCESS** The Open Access icon, which is a stylized padlock with an open keyhole.