



The automation of Clinical Data Management Driven Activities

Version 1 (June 2022)

Authors

- Demetris Zambas, VP, Global Head, Data Monitoring & Management, Pfizer
- Lynne Cesario, Sr. Director, Global Risk Based Monitoring Program Lead, Pfizer
- Patrick Nadolny, Global head, Clinical Data Management, Sanofi
- Prasanna Rao, Sr. Director, Artificial Intelligence and Machine Learning, Pfizer

Reviewers and Contributors

- Sanjay Bhardwaj, Head of Clinical Technology Strategy & Operations, Abbvie
- Jérémy Nadolny, Sr Business Associate, Trial Operations

SCDM would like to acknowledge the many volunteers who have contributed to development of the SCDM reflection papers used as the primary basis for this topic brief.

Methodology

The Society for Clinical Data Management (SCDM) Innovation Committee seeks to provide Thought-Leadership to our industry and support the SCDM vision of “*leading innovative clinical data science to advance global health research and development*”. To that end, the SCDM Innovation Committee strives to demystify Clinical Data Science (CDS) and support the development of all Clinical Data Management (CDM) professionals, from subject matter experts (SMEs) working on clinical studies to CDM leaders setting the direction of their organizations.

The Innovation Committee is publishing topic briefs intended to serve as orientation guides on specific topics which are contributing directly or indirectly to the evolution of CDM into CDS. The content of those topic briefs is primarily an extract from the previously published SCDM Reflection Papers^{1,2,3} which collectively provide a cohesive and comprehensive overview of CDS from the point of view of industry leaders. Due to the recent emergence of the CDS discipline and the absence of a comprehensive literature base regarding CDS within the Drug Development industry, this content was gathered from industry leaders through a consensus-based methodology. As CDS mature, and technology evolves, it is anticipated that literature on this topic will blossom.

Introduction

The use of “Artificially Intelligent” applications is fast growing outside our industry. These artificial intelligence (AI) “based technologies have the potential to transform healthcare by deriving new and important insights from the vast amount of data generated during the delivery of healthcare every day”⁴. Intelligent applications have long left the Silicon Valley computer labs and have entered many homes. Interactive systems like Alexa, Apple Siri and Google Home have gained adoption and trust. This evolution is primarily fueled by the accumulation of huge amounts of data associated with strong computing power. This is enabling the training and validation of complex intelligent solutions which will revolutionize clinical development and dramatically change CDM at its core.

Topic brief



This topic brief will intentionally focus on automation technologies as these are mature and their applicability to CDM is immediate. While the automation of many CDM activities is unavoidable, the future of the CDM profession has never been so bright and so central. In a world where people and intelligent machines will collaborate, CDM leaders can leverage their deep data knowledge to take proactive steps and articulate a clear capability strategy toward CDS that considers all aspects: people, process, regulations, standards, and intelligent technologies. But unfortunately, when it comes to AI, many organizations do not know where to start or waste their energy going in many directions without a clear vision. They may not even fully understand what AI is and where to leverage it⁵. For those in search of a path forward, AI based automations of CDM driven activities may well be a first meaningful and impactful move. In some cases, CDM could be an enabler for the automation of activities owned by other functions in Clinical Development (e.g., detection of data driven Protocol Deviation).

Automation technologies

First and foremost, organizations must anticipate the implication of using AI based solutions leading to the set-up of digital workforces that can work 24 hours a day and be “hired” (i.e., implemented) and on-boarded quickly and at a decreasing cost. This digital evolution will profoundly transform the workplace in years to come. Also, as we leverage advanced capabilities, we need to augment our approaches to their testing, deployment, and management. Adoption and change management of any type of automation requires SME participation from the inception to delivery of the solution.

Intelligent technologies driving automation

Before diving into automation, it is important to clarify the underlying technologies enabling the development of “artificially intelligent” applications and the differences between AI, Robotic Process Automation (RPA), Intelligent Process Automation (IPA), and other related technologies. Figure 1 shows the key components necessary to power automations. These components have different roles to play and could be used either individually or combined to deliver the desired CDM automations. They start with process automation (i.e., RPA) and evolve toward the simulation of human cognitive behaviors such as reasoning and learning by machines (i.e., true AI). They collectively enable a wide range of opportunities for CDM that will impactfully transform our discipline toward CDS. This transformation also empowers Clinical Data Scientists to further enable the leveraging of such automations across the clinical development ecosystem.

Robotic Process Automation (RPA)	Intelligent Process Automation (IPA)	Natural Language Processing (NLP)	Natural Language Generation (NLG)	Machine Learning (ML)	Artificial Intelligence (AI)
Basic Automation of repetitive and predictable tasks	Advanced Automation of non-routine tasks	Language and text analysis	Generate text and voice	Pattern recognition	Human like behaviors Self optimization and Learning
No / simple decision making	Some decision making and problem solving	Knowledge gathering	Knowledge sharing	Prescriptive and predictive analysis	Deductive analysis

Fig 1. Technologies powering automations

First, **AI** is an evolving field and umbrella term which typically includes natural language processing (NLP), natural language generation (NLG) and machine learning (ML). It also includes IPA which is the natural evolution of RPA. With IPA, robots can perform non-routine tasks requiring more complex problem solving.

RPA is process driven, meaning it focuses on process automation, whereas **AI**, which focuses on thinking and learning, **is data driven**. RPA is considered by some to be the simplest form of AI. The fact is that RPA solutions do not require problem solving capabilities and are at best using prescriptive decision trees. Scripting historically done by Clinical Data Managers are in many ways similar to RPA except that the ‘script’ in of itself did not have a user profile like a bot would have, such as in the case of RPA where a bot can login to many source systems to perform automation of tasks such as acquire and aggregate data to deliver a desired output (e.g., SAE reconciliation).

IPA combines both RPA and components of AI. IPA is often powered by ML, NLP and NLG which were covered in the reflection paper Part 1. Smart workflows monitor and track information exchanged between people and systems in real time. Then, ML can use pattern recognition to support intelligent decision making while NLP can convert unstructured content to be interpreted (e.g., to enable RPA bot-based workflow).

NLP and **NLG** are used by machines to process human language to extract specific information from documents, to generate text from voice (or vice versa), to generate parameterized text (e.g., NLG can be used to create a Query text that was traditionally written by a data manager, understand meaning, etc.). Today, machines can accurately carry out simple requests, ease web searches, summarize documents or translate languages. NLP is one of the key enabling technologies behind chatbots and spam email detection.

While NLP is almost 70 years old, it has tremendously progressed over the last decade by adopting machine learning techniques instead of solely trying to automate grammar rules typically full of exceptions. Apple Siri released in 2010 and Amazon Echo (i.e., Alexa) released in 2015 are two examples demonstrating this technological breakthrough, which are now part of the daily routines of many people across the globe and can manage many different languages. Looking beyond these “simple” use cases, NLP can offer many opportunities to clinical development and drug safety processes within biopharma companies.

NLP can be leveraged to “*extract the names of drugs, diseases, patients, and pharma companies using rule-based or statistical method*”⁶ in order to support Pharmacovigilance activities as a vast amount of information about Adverse Events (AEs) reside in unstructured narratives. Similar methodologies could be applied to **identify potential AEs from comments** captured in EDC, eCOA or other systems which is currently

labor intensive and subject to human errors. NLG could then be used to automatically generate patient's summaries once all relevant information has been extracted through NLP from raw data.

Last but not least, **Knowledge Graphs** can add context and depth to AI Based methodologies. The term 'knowledge graph' was introduced by Google in 2012 to refer to its general-purpose knowledge base. A knowledge graph organizes and integrates data according to an ontology, which is called the schema of the knowledge graph, and applies a reasoner to derive new knowledge. Knowledge graphs can be created from scratch (e.g., by domain experts, learned from unstructured or semi-structured data sources) or assembled from existing knowledge graphs, typically aided by various semi-automatic or automated data validation and integration mechanisms⁷. In the field of CDM, Knowledge Graphs can be created from documents such as Protocols, Statistical Analysis Plans, Clinical Summary Reports along with other specification documents like Study Design Specification and Metadata. Knowledge graphs can link, automate, trace and optimize data collection.

Robotic Process Automation

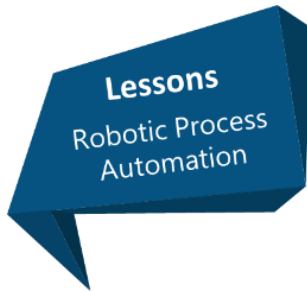
RPA is the simplest form of automation technology which is implemented through software applications, or bots, that perform predictable and routine tasks that do not require complex decision making. RPA is a mature technology which has been widely used across industries for years. It can mimic human interactions across multiple applications and is **usually implemented to improve efficiency without the need to deeply change existing processes**. Implementing RPAs helps reduce process time, increases throughput, enhances accuracy and frees up human employees by helping them to focus on the tasks that bring the most value back to the organization. Typically, RPA applications take the form of a "bot", which can be thought of as digital employees (i.e., Virtual Clinical Data Manager). As an example, a bot could be fed from the outcome of external data reconciliation errors. The bot could then login to the relevant system with its own account (e.g., Login CDM-Bot-007) and take the appropriate action defined by the automation (i.e., edit, copy/paste, download, upload, etc.). The advantage of such a method is that it does not require changes to the system (i.e., no integration needed). It leverages the system's native traceability (Login logs, Audit Trail, etc.).

RPA scripts using simple branching logics and Boolean operators (e.g., and, or, not, etc.) can be applied to automate repetitive and transactional processes. Using those scripts, an RPA bot will act very much like an end user, capable of logging in and out of systems to perform basic process-oriented tasks using well defined data inputs. One needs to understand that RPA requires programming, full understanding of the domain being automated and is limited in its scope of applicability. Additionally, as RPAs are sensitive to User Interfaces (UI) layouts (i.e., RPAs must be adjusted if the UI is changed), CDM need to consider the stability of the systems involved in task automation or anticipate updates to RPAs when systems are upgraded. So, new processes and procedures are required to test and manage the identity and deployment of bots, given they will perform steps like any human, and their actions will be recorded in systems audit trails. Auditors and inspectors will undoubtedly seek to understand how RPA bots were developed, tested, and supervised by humans in production environments.

Clinical Data Scientists must not only be able to identify the appropriate use cases where RPA can be applied but be able to assess the value of automating such use cases given the validation requirements and time and costs involved. Both a positive return on investment (ROI) as well as a means of tracking value (quantitatively as well as qualitatively) are critical for organizational support. The cost needs to be balanced

with quality as the automation of large scale repetitive manual tasks could result in meaningful process accuracy and reliability improvements reducing risks.

To unlock its value, it is imperative to begin by setting a framework to educate all clinical research stakeholders and evaluate prospective opportunities for automation. Below are five best practices stemming from the lessons learned from early adopters that could be leveraged when implementing RPA.



Best Practice #1: Establish and gain cross-functional alignment on an Automation charter and deploy a focused supporting team. First, it is advisable for organizations to establish a dedicated team that is responsible for meeting the organization’s vision and delivering its high-level strategic roadmap. The team should be cross-functional, with representation from all roles supporting clinical studies as well as enabling functions (e.g., IT, finance, etc.). This team should be chartered and provided the means to deliver impactful business automations. Each automation must set clear objectives and measurable outcomes (e.g., cycle times / resources reduction, quality / accuracy increase by X%, etc.).

Best Practice #2: Educate the team on AI principles and their scope of applicability. As previously mentioned, AI is a combination of capabilities that allow machines to replicate human decision-making and/or interactions. Each capability offers different opportunities and is associated with different levels of complexity. Once these different principles are clear, it is easier to strategize and focus on specific RPA opportunities.

Best Practice #3: Inventory potential automatable processes. There are many ways to collate ideas, however, onsite workshops focused within each business line can accelerate the process by bringing together functional SMEs to review ideas and opportunities in real time, and provide a forum for others to discuss and build on them. Each SME should be prepared to present a description of the automation and where it fits in the high-level process, including any potential changes that may be required to the process to enable the automation. For instance, data may need to be stored in a different location to enable the bot, or access to a data source might need to be opened to a different business group that was historically not considered. The business value should also be clearly articulated. The expected outcome of these sessions is not only to identify opportunities for automation, but also to prioritize the opportunities for delivery and estimate potential time and/or cost efficiencies and cycle-time reductions.

Best Practice #4: Establish a clear automation roadmap. To ensure successful RPA implementation, it is important to establish a clear short-term and long-term automation roadmap. After assembling and reviewing all of the automation use cases, the team should prioritize the projects with the highest alignment with the company’s strategic goals, and use those to set the foundation for how to implement AI/RPA within your department. As CDM review each idea, they can clearly define which components of the AI principles would be utilized to roll out the solution. Projects only requiring RPA capabilities, rather than those requiring hybrid solutions, can be prioritized for easy and rapid implementation. The team should also align on a metric that will be reported out once the automation goes live, training/change management is in place and an adequate “ramp up” window is accounted for to ensure full adoption of the new process. Once the organization has started its RPA journey, it is important to track utilization of each bot to ensure the full value of the automation is truly realized. Low bot utilization could also point to issues

with how it was programmed or implemented, and that should be addressed via programmed updates over time to ensure that the bots grow with the business process changes.

Best Practice #5: Build a sustainable automation development process and governance. A well governed and repeatable process is necessary to secure RPA operationalization. This means developing a process with input from all relevant stakeholders, testing the process and validating each bot to ensure the workflow is correct and efficient. Organizations need to consider establishing AI governance to build transparency and trust⁸. We should also consider ethical aspects, including bias that could potentially be introduced from data.

Potential areas of RPA automation to be assessed by the governance team could include:

- Serious adverse event reconciliation
- Automated clean patient tracking
- Auto-generation of edit check specifications
- Automated loading of 3rd party data
- Automation of clinical data reviews
- Analysis and reporting system process triggers (scheduler)
- Automated quality review

AI based Smart Data Query

One large sponsor organization conducted a ‘hackathon’ which focused on the pattern recognition through Machine Learning of the query management process. It involved the review of thousands of clinical data points to determine if a data discrepancy should be queried with a site. Many studies have thousands of manual queries raised over a long period of time (lasting from several months to many years). Besides prescribed automated “edit checks”, CDM must reconcile data across several CRFs and data sources by reviewing data and audit trails manually to determine if a query must be raised or not.



The hypothesis of the hackathon was that smart machines can learn from historical data by associating multiple clinical datapoints to specific manual queries, and establish clear data to query association patterns. When a similar pattern is presented to this trained model, the expectation was that it could predict discrepancies and, possibly, raise queries.

Data from several historical studies from various EDC systems were provided to the hackathon participants in a secure cloud environment. The studies had multiple data standards and hundreds of millions of clinical data points. ML models were developed by four companies using various state-of-the-art techniques including deep learning. Vendors were evaluated on various aspects including innovation, accuracy and their ability to scale to a production solution. The outcome was successful in establishing a method for intelligent machines to perform CDM activities and proved that the hypothesis was correct. A semi-supervised learning approach was determined as the most successful method. Some labeled data were provided for supervised machine training by CDM SMEs and combined with an unsupervised learning algorithm where machines predicted similar patterns through observations.

As a next step beyond the hackathon, the selected ML was trained with Sponsor organization's data beyond those provided for the hackathon to ensure strong accuracy and consistency of the predictions for critical data domains such as adverse events (AEs) and concomitant medications (CMs). Once perfected, the model was implemented in production by integrating it directly into EDC and intelligent CDMS systems. At this stage, these models can reliably detect complex cross-panel data discrepancies and save valuable time. The 'human in the loop' method enables CDM to review all predictions, determine if they are correct, and automate the generation of queries.

In future, NLG may ultimately mature to the point where machines can compose queries depending on the discrepancy context. The approach of using historical discrepancies and their associated queries will only train machines on known patterns. Future work to expand this use case to introduce clinical reasonings could move the AI needle more toward CDS. For example, could machines detect AEs that are not reported? Can we take a subject holistically, analyze and reason whether a concomitant medication should have been taken when there was a reported AE, etc.

Overtime, the model learnings from other clinical studies and SMEs will expand the horizons of AI and CDS. The technology can look across TAs, identify more complex patterns and adapt over time, while humans will assess machine inferences and provide invaluable feedback. In this context, computers and humans working together will make better predictions than either could do on their own. Additionally, technology that never stop learning and improving generates endless opportunities for CDS, the processes they support and the organizations they serve.

Lastly, to successfully unleash the potential of AI-based automations, people will need to validate that this method is reliable, and that the RPA or IPA based bot is a dependable 'digital assistant' that interacts with employees and partners to make important decisions. If an AI application is not well designed and managed, it may 'misbehave', with significant quality and ethical ramifications. These could range from damaged site relationships to negative impacts on patient safety and trial reliability that could affect quality and discriminatory decisions that elicit regulatory scrutiny.

Considerations for the validation of ML based solutions

First, it is very important to understand that ML is designed to make predictions or decisions from data without explicit programming. Like humans, ML Tools will make decisions tomorrow that are better than the one they made today. The ability for "*ML software to learn from real-world feedback (training) and improve its performance (adaptation) makes these technologies uniquely situated among software*"⁴. So, CDM needs to carefully anticipate the implication of evolving systems in our regulated environment when reproducibility of results is expected. This has prompted FDA "*to reimagine an approach to premarket review for AI/ML-driven software modifications*"⁴. While focused on Software as Medical Device (SaMD), this FDA discussion paper highlights the need for new approaches with AI and for Good Machine Learning Practices (GMLP). We also need to realize that ML learning could be biased by the training datasets or if human supervisors are 'just' accepting the decisions from the solution failing to provide adequate safety nets which would re-enforce the biased behavior.

Validation of new ML Based solutions goes beyond the healthcare industry and has been extensively discussed in the literature⁹. GMLP recommend the use of three independent datasets in the development

and validation of a new ML based solution. A model is trained on a specific dataset called the **training set** and is further optimized on an independent dataset, the **development set**. This optimization is an iterative process where the model parameters are tuned at each iteration to maximize specific performance criteria. This step is critical to ensure that models can generalize to new data while maintaining sufficiently high levels of performance. Finally, the performance of the optimized model is assessed on a third dataset, **the test set**. This step ensures that the validation of the model is performed independently from the model building process and that performances are not over-estimated.

Stakeholders need to play a key role in the definition of performance criteria and need to be trained in properly interpreting those metrics. This will leverage their power in making informed decisions regarding the release of new ML based solutions in a production environment. As an example, the accuracy, i.e., the ratio of the number of correct predictions to the total number of input samples, is a common performance criterion used to optimize ML solutions. However, the optimization of this unique criterion may fail to produce models that are relevant for specific clinical use cases. Let us consider the optimization of a ML based solution, trained on the overall US population, to detect patients that will develop a rare disease. The US population is an unbalanced training set where the number of subjects with the disease (positive cases) is much lower than the number of subjects without the disease (negative cases). The optimization of the accuracy criterion alone in that setting would likely favor the detection of negative cases at the cost of missing true positive cases. In clinical development and overall healthcare settings, “sensitivity” and “accuracy” have distinct meanings. When targeting detection of critical issues in a clinical trial it may be preferred to target slightly higher false positives to minimize the possibility of false negatives.

Once the model is released to production, several factors can lead to the degradation of its performance over time including a modification of the user’s behavior or to the input data. A monitoring process (i.e., a human in the loop) should be put in place to actively assess the model performance over time and trigger specific actions when the model performance drops below a predefined threshold. These actions include the training and optimization of the model on new datasets and adaptation of the model architecture.

Considerations for the deployment of Intelligent Solutions

Intelligent solutions powered by AI require the most significant process, technology and mindset transformations for CDM. Overall, ML based solutions will act as virtual Clinical Data Managers assisting Clinical Data Scientists. As a result, expert Clinical Data Scientists will mentor virtual Clinical Data Managers to accurately perform data reviews and other CDS activities. This means that Clinical Data Scientists will need to define the objectives of the intelligent solution and identify the datasets required for training and testing. These datasets will need to cover most expected data review scenarios (i.e., be complete) and be truly representative of the use cases anticipated in production (e.g., include data review scenarios across all study phases and therapeutic areas).

The failure to define the right datasets could bias the system behaviors and lead to inconsistent data review accuracy. Clinical Data Scientists also need to play the key role in assessing the performance of the solution by defining the right testing strategies (e.g., by setting a minimum ratio of the number of correct predictions compared to the total number of inputs in the test dataset).

The figure below is an example of a process for the “mentoring” (i.e., training) of an ML-based solution.

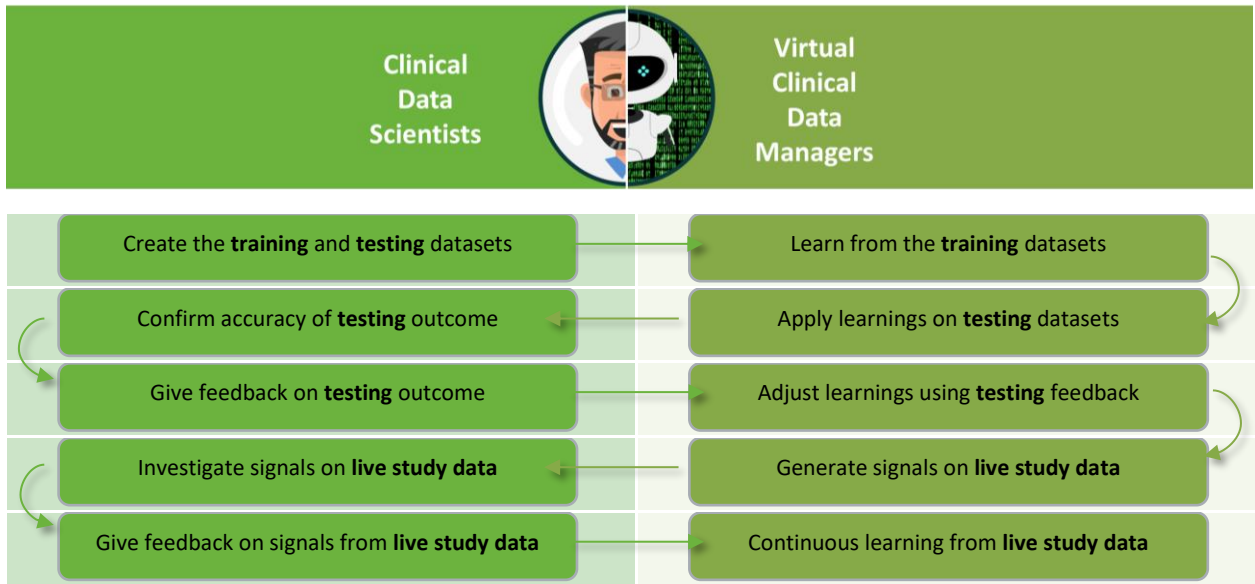


Fig 2. Example of learning process for an ML-Based solution

ML technologies require fundamentally different approaches to their development and validation including the use of methods like recall (i.e., ability to predict in all cases), precision (i.e., accuracy of the predictions) and F-measures (i.e., combining recall and precision) to determine the reliability of models.

First, the model must be trained on a controlled dataset of known quality. Then, it can be applied to the testing dataset to assess the precision and recall in a simulated “real-world” environment to understand its precision and recall.

Its targeted level of accuracy should be one that is better than the accuracy of the current process assuming the current process performance is known and documented. However the acceptable initial implementation accuracy may be lower. If initially lower, which can be common, its use needs to include commensurate human supervision and training to monitor its accuracy. Lastly, once a machine learning model is deployed, it may need to be retrained at specific intervals in order to maintain and ultimately improve its accuracy over time. The retraining will need to be versioned controlled to ensure full lineage for reproducibility, auditability, and inspection readiness.

CDS organizations will need to leverage its risk-based data management strategy when implementing such disruptive technology to ensure reliable and ethical (i.e., unbiased) decision making. In the early adoption steps, the level of supervision (i.e., the human in the loop) is crucial to safeguard from any impact on patient safety or the reliability of the trials results. The supervision measures need to be commensurate to the assessed risk. To one extreme, solely using an NLP solution to review of source data to detected unreported AEs would represent a high risk and should strongly be discouraged. However, as highlighted in our Smart Data Query use case, piloting the automatic generation of queries in addition to existing CDM processes (e.g., Manual review and edit checks) would not introduce any risk and could be rapidly deployed to independently identify discrepancies in the non-critical data.

Conclusion

There are numerous opportunities for technology-enabled Intelligence combining “humans” and “machines”. Augmented Intelligence is often the preferred term rather than Artificial Intelligence. We could foresee a Clinical Data Management world built on a *working model that includes virtual Clinical Data Managers working alongside Human Clinical Data Scientists*. AI could enable humans to achieve a different level of reasoning and performance not otherwise possible. CDS organizations need to develop and provide internal frameworks which enable the development of capabilities to automate repetitive CDM tasks and thereby leaving the human Clinical Data Scientist to focus on higher value critical-thinking tasks including the monitoring of the performance of the AI-enabled automation, reduce bias, apply risk-based quality approaches to problem solve etc. By doing this, computers and humans working together could make better predictions than either group of humans or an assembly of machines could do on their own and therefore revolutionize drug development.

Main abbreviations

AI	Artificial Intelligence
CDM	Clinical Data Management
CDS	Clinical Data Science
DCT	Decentralized Clinical Trials
eCOA	Electronic Clinical Outcome Assessments
EDC	Electronic Data Capture
GMLP	Good Machine Learning Practices
IPA	Intelligent Process Automation
ML	Machine Learning
MHRA	Medicines and Healthcare products Regulatory Agency
NLG	Natural Language Generation
NLP	Natural Language Processing
ODM	Operational Data Model
RPA	Robotic Process Automation
RWD	Real World Data
ROI	Return On Investment
RWE	Real World Evidence
SCDM	Society for Clinical Data Management
SaMD	Software as a Medical Device
SME	Subject Matter Expert
UI	User Interface

References

- ¹ SCDM, June 2019, The Evolution of CDM to Clinical Data Science – A Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management. Available at <https://scdm.org/wp-content/uploads/2019/09/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science.pdf>
- ² SCDM, March 2020, The Evolution of CDM to Clinical Data Science (Part 2: The technology enablers) – A Reflection Paper on how technology will enable the evolution of Clinical Data Management to Clinical Data Science. Available at <https://scdm.org/wp-content/uploads/2020/03/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science-Part-2.pdf>
- ³ SCDM, August 2021, The Evolution of CDM to Clinical Data Science (Part 3: Evolution of the CDM role) – A Reflection Paper on the evolution of CDM skillsets and competencies. Available at <https://scdm.org/wp-content/uploads/2020/08/SCDM-Reflection-Paper-CDM-Role-Evolution-Part-3-.pdf>
- ⁴ FDA, 2019, Discussion and Request for Feedback, “Proposed Regulatory Framework for Modifications to Artificial/Machine Learning [AI/ML]- Based Software as Medical Device (SaMD). Available at <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>
- ⁵ Patrick Nadolny, Clinical Leader, September 2020, “What Is Your AI Road Map To Revolutionize Drug Development?”. Available at <https://www.clinicalleader.com/doc/what-is-your-ai-road-map-to-revolutionize-drug-development-0001>
- ⁶ Marco Lagi, Natural Language Processing – Business Applications, February 2019. Available at <https://emerj.com/ai-sector-overviews/natural-language-processing-business-applications/>
- ⁷ The Aman Turing Institute, Knowledge graphs - How do we encode knowledge to use at scale in open, evolving, decentralised systems? Available at <https://www.turing.ac.uk/research/interest-groups/knowledge-graphs>
- ⁸ Cognizant, Jan 2019, Making AI responsible and effective, Available at <https://www.cognizant.com/whitepapers/making-ai-responsible-and-effective-codex3974.pdf>
- ⁹ Andrew Ng, 2018, Machine Learning Yearning, available at <https://www.deeplearning.ai/machine-learning-yearning>