Paper ET01

A Data Manager’s Experience with AI/ML Adoption and Efficiency

Diane Lacroix, eClinical Solutions, Mansfield, MA, United States

ABSTRACT

AI buzz continues to include exciting predictions along with questions about job security, regulatory guidance and oversight. But AI intent is here in clinical development, and practical use cases are being implemented today with the potential to deliver much-needed efficiency gains. Successful AI adoption requires focusing on the users of these AI algorithms – humans providing necessary supervision and reducing risk, but also fine-tuning models, actioning decisions and finding new, efficient ways to work smarter using this intelligence. This session will examine the data manager’s experience with AI/ML from the practitioner perspective. With a focus on AI-enabled data review, the talk will highlight clinically relevant AI use cases for DM and address how data management uses these outputs to augment their data review and adapt typical workflows. Learn how as result, the role adapts to incorporate new skills in data science and technology as they inform, adopt and utilize artificial intelligence applications.

INTRODUCTION

AI and ML are exciting topics! Before we dive in, let's start with a quick terminology refresher.

First, what is Artificial Intelligence or AI? To paraphrase the definition - AI refers to the simulation of human intelligence in machines that are programmed to think, reason, learn, and problem-solve like a human.

Second, Machine Learning, or ML, is a subset of AI that focuses on the development of algorithms and statistical models that enable computers to learn and make predictions or decisions without being explicitly programmed. In essence, it's about training machines to improve their performance on a specific task through data-driven experiences.

It's important to note that while AI is the broader concept, ML is a key component of AI. In other words, all Machine Learning is AI, but not all AI is necessarily Machine Learning.

DRIVING THE EVOLUTION

What is driving the need to transform and evolve data management approaches within the life sciences?

- Trial Complexity: Master Protocols which include:
  - Basket: Multiple Diseases One Drug
  - Umbrella: One Disease Multiple Interventions
  - Platform: a type of master protocol that tests multiple, targeted therapies for a single disease simultaneously. Platform protocols often include an adaptive design that may eliminate or add treatments based on interim analysis.

The Amount of DATA! This is not new news that the sheer volume of data being collected, the variety (labs, wearables, genomics, ePRO, eCOA, Decentralized Trial Technologies), and the velocity at which this incoming data is streaming into trials, is continually increasing.

This highlights the need for AI/ML in Clinical Data Management data review to help ensure accuracy and data quality for decision-making at the accelerated pace that is needed. There is a need to evolve and change the approach, which includes adoption of risk-based strategies and automation of repetitive and routine tasks to reduce manual efforts and free up data management and study teams to focus on critical data that matters. Finally, we must embrace AI tools as virtual data managers to surface trends and data issues, propose queries, and act as partners to inform human data managers to make the right decisions and take appropriate action.

DEMYSTIFYING THE HYPE

Let’s start to demystify some of the hype and look at the data manager’s responsibilities in a traditional approach to data review as compared to an AI-augmented approach to discern differences and similarities.

In the **Traditional Data Review Model**, there are a subset of reviews that are complex, are difficult to program efficiently, and require significant manual time and effort to review. As manual review is performed, the Data Manager is making decisions, determining actions, and raising manual queries to sites or perhaps data issues to their teams.

- Manual review is time-consuming and therefore resource intensive.
- Human reviewers may interpret data differently or make mistakes when reviewing large amounts of data, resulting in inconsistencies and errors that can negatively impact data quality.

With **AI Augmented Data Review**, the model is doing the bulk of the processing and surfacing to the human (the data manager) focused outputs such as anomalous data or patterns, data discrepancies and outliers. The model reduces or eliminates the need for review of the data that the human does not need to look at, based on what the model understands.

The CDM serves as the "human in the loop," reviewing the results of the model output and making the final decisions on how to action the data issue.

- AI/ML-powered approaches can save time and reduce the risk of errors or oversights in data review.
- CDMs are more productive as they do not waste time on non-value-added tasks and can apply their domain expertise to more critical tasks.

While there are clearly differences in these approaches, there are also similarities in that there will still be the need for the human, the data manager, to review the model output, and make the final decision on how a data issue will be actioned. The reality is, it is not that different except for the adoption of smarter and more efficient tools that help rather than hinder.

HUMAN IN THE LOOP (HITL)

The human in the loop is critical. Here is why. 'Human in the Loop' is a type of AI that leverages both human intellect and machine intelligence to create machine learning models.

1. Training Data is INPUT into the model to help the model understand the world or in the case of data review, to understand the review objective it is intended to meet.
2. The machine learns to make decisions from this training data
3. For AI models where predictions fall below a certain level of confidence, the Human in the Loop gives direct feedback into the model
4. Because of the human's feedback, the machine keeps adjusting its "view of the world" so that the algorithm gets smarter, more confident and more accurate.

Human-in-the-Loop aims to achieve what neither a human being nor a machine can achieve on their own. When the machine is not able to solve a problem, humans need to step in and intervene. This process results in the creation of a continuous feedback loop. With constant feedback, the algorithm learns and produces better results every time under this human supervision.

DATA MANAGEMENT (DM) AS HUMAN IN THE LOOP (HITL) - DEVELOPMENT

A key role of Data Managers in the HITL AI workflow is to act as the domain expert in the AI development process. The AI development process is an iterative process where teams strive for continuous maintenance and improvement. It starts with trying to understand what the business needs are and assess the value of automation versus the underlying cost. Next is understanding the available data and doing exploratory data analysis before building the model and evaluating its performance. If the performance meets the criteria, the model can be deployed. After its deployment the model is monitored for performance and maintained, that is, check for changes in the behavior of the data or the underlying relationship between the inputs and outputs and update the model if needed. The model would also be updated when user feedback is received. The process is cyclic because of the continuous cycle of problem definition, model development and deployment and monitoring process.

There are many ways that the data manager helps throughout this model development cycle. It starts with defining the business need. Data managers help identify pain points and discuss opportunities where AI solutions may be applicable. As data is gathered and preprocessing performed, data management domain experts help provide understanding of the data, for example, how it was collected and what errors exist and guide the selection of adequate data - data appropriate for the intended use case and data applicable for the intended target population. The data needs to be complete and representative of real-world cases. Another important task is that data managers provide domain expertise in feature extraction, which is the creation of the inputs that are fed into the model. This process of data collection and preprocessing is critical to the success of the project because better and more data leads to better solutions.

One task that relies heavily on domain experts is data labeling. In ML, there are two main types: supervised learning and unsupervised learning. In supervised learning, labeled data is used to build the AI models. Labeled data means that we know what the true answer should be, and this is the answer that we want the model to predict. And the human experts provide those true answers to us. In unsupervised learning, the AI model is built without labeled data. In this case we do not know what the true answer is. Even if using unsupervised learning methods, having domain experts label some portion of the data helps check performance.

Once we have the labeled data, we move on to model training. The role of the DM during this phase is to help answer any questions we have about model choices and fine tuning of parameters.

After training, the domain expert plays a large role in model validation. For model validation, it's important to understand what the measure of success is because there's no one best metric. You may have heard of some of these metrics: sensitivity, specificity, precision, recall, F1, negative predictive value, positive predictive value, ROC (receiver operating characteristic curve). The best metric or metrics depends on the data itself and what we're trying to achieve and how the model will be used. And it's a balance between these metrics. For example, you can only get better sensitivity by sacrificing specificity. It’s important to work closely with the data manager to understand how best to measure success because we need to consider how the data manager will be using the model. The data manager also helps to validate and interpret the results to make sure that they make sense and that the model is impactful. Another reason that we want to work closely with data management in evaluation is that knowledge of model performance helps the human to interpret the results and make informed decisions when they start to use this Virtual data manager in their work. Finally, we ask that the data manager sanity checks the model results to make sure that they make sense and that they are impactful.

Based on the validation results, the go/no-go decision is made together on deployment. Once the model has been deployed, we move onto monitoring phase we continually check to see that the model is performing as expected.

FOUNDATION FOR AI/ML

AI on its own will not solve all problems, but it can help us solve problems faster and more efficiently. That cannot happen without the right foundation. As a foundation for an AI ML initiative, there first needs to be a strong technology infrastructure in place, that is capable of ingesting all of the trial data to organize, transform and aggregate to one central source. This technology foundation should be capable of ingesting data from any source in near real time, transforming that data to a central source of truth, and powering data insights across data sources through the use of different modules such as Management Workbenches and Analytics modules.

CASE STUDY: TRADITIONAL DATA REVIEW VS AI-AUGMENTED DATA REVIEW

AI models are clearly reducing effort by bulk processing data and surfacing data that requires human review and intervention. This case study illustrates the point.

In this case study, in the traditional approach to data review, for 55000 records, reviewing output manually at 8 min/record amounts to 7333 hours.

In the AI Augmented Review approach, using four AI models where the models are doing bulk processing of the data, and surfacing only what requires the human in the loop to review, resulted in a 70 – 90% reduction in the records to review.

Records to review went from 55000 to 16500 records with 70% reduction and to 5500 records with the 90% reduction. Translating to hours, we went from 7333 to between 2000 to 733 hours.

Continuing to identify opportunities to automate data review, we confidently expect to see further reductions across the data review lifecycle, improving patient safety and facilitating higher quality data for decision making, sooner in the process.

CONCLUSION AND KEY TAKEAWAYS

AI ML flags data anomalies and inconsistencies, allowing data managers to focus their efforts on resolving specific issues vs manually reviewing data. However, AI ML can only succeed with human professionals' supervision, expertise, and contribution. For this reason, the data manager is a critical Human in the Loop. The data management skill set, which includes domain expertise, adaptability, communication, ethics, quality assurance, collaboration, feedback, documentation, and training are key to transitioning to Clinical Data Science.

AI ML REDUCES EFFORT

Use of AI ML models and the efficiencies gained allow data managers to be more efficient and productive to focus their efforts on more critical tasks vs manually reviewing data.

HUMAN EXPERTISE IS ESSENTIAL

Human Intellect and Machine Intelligence together are needed for AI ML models to succeed.

DATA MANAGERS ARE POSITIONED FOR SUCCESS

Data Managers understand data. They are the data stewards and know the data better than anyone. The variety of skills data managers currently need to have to be successful today position them to be successful in adopting these newer technologies to support the evolution towards clinical data science.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Diane Lacroix
eClinical Solutions
603 West Street
Mansfield, MA, United States 02048
Email: dlacroix@eclinicalsol.com
Web: www.eclinicalsol.com