PERFICIENT®

How Artificial Intelligence Can Enhance the Clinical Data Review and Cleaning Process

Ensuring that clinical trial data is accurate and that clinical trials are safe and effective, is a time-consuming and a manually intensive process. Many life sciences companies have implemented home-grown or off-the-shelf clinical data review platforms and defined the reviewing and cleaning processes to be used with them. Many of these systems have proven to be challenging to use, inflexible, and created frustration among users.

This guide discusses how artificial intelligence (AI) – including machine learning (ML), deep learning (DL) and natural language processing (NLP) – can be used by pharmaceutical and medical device companies to improve the clinical data review and cleaning process. These technologies enable drugs and devices to reach the market faster and more safely and effectively.

Human-Computer Systems

Humans and machines each have their own strengths. On the one hand, machines are good at processing and analyzing large volumes of data with high speed and accuracy. On the other hand, humans are good at making decisions based on data, interacting with other humans, and applying general intelligence to the data. When it comes to reviewing clinical data, a human-computer system will perform better than either standalone method. The AI initiatives that revolve around clinical trials require humans and machines to work together.

Machine learning models can understand clinical data, do an initial analysis of the data, and perform an initial cleaning of the data using statistical analysis. Based on the clean/dirty probability, humans (data reviewers) can prioritize their review activity. They can search data using NLP and obtain the status of their activities using a self-documenting platform that gives them a current summary of the clinical trial state. Analysis using ML models can provide more insight into clinical data and enable humans to determine the safety and efficacy of the trial.

Analysis using ML models can provide more insight into clinical data and enable humans to determine the safety and efficacy of the trial.

Clinical Data Review Platforms (CDRP)

While ML and NLP functionalities are not currently available in today's CDRP's, we know they offer tremendous benefits to the clinical data review and cleaning process.

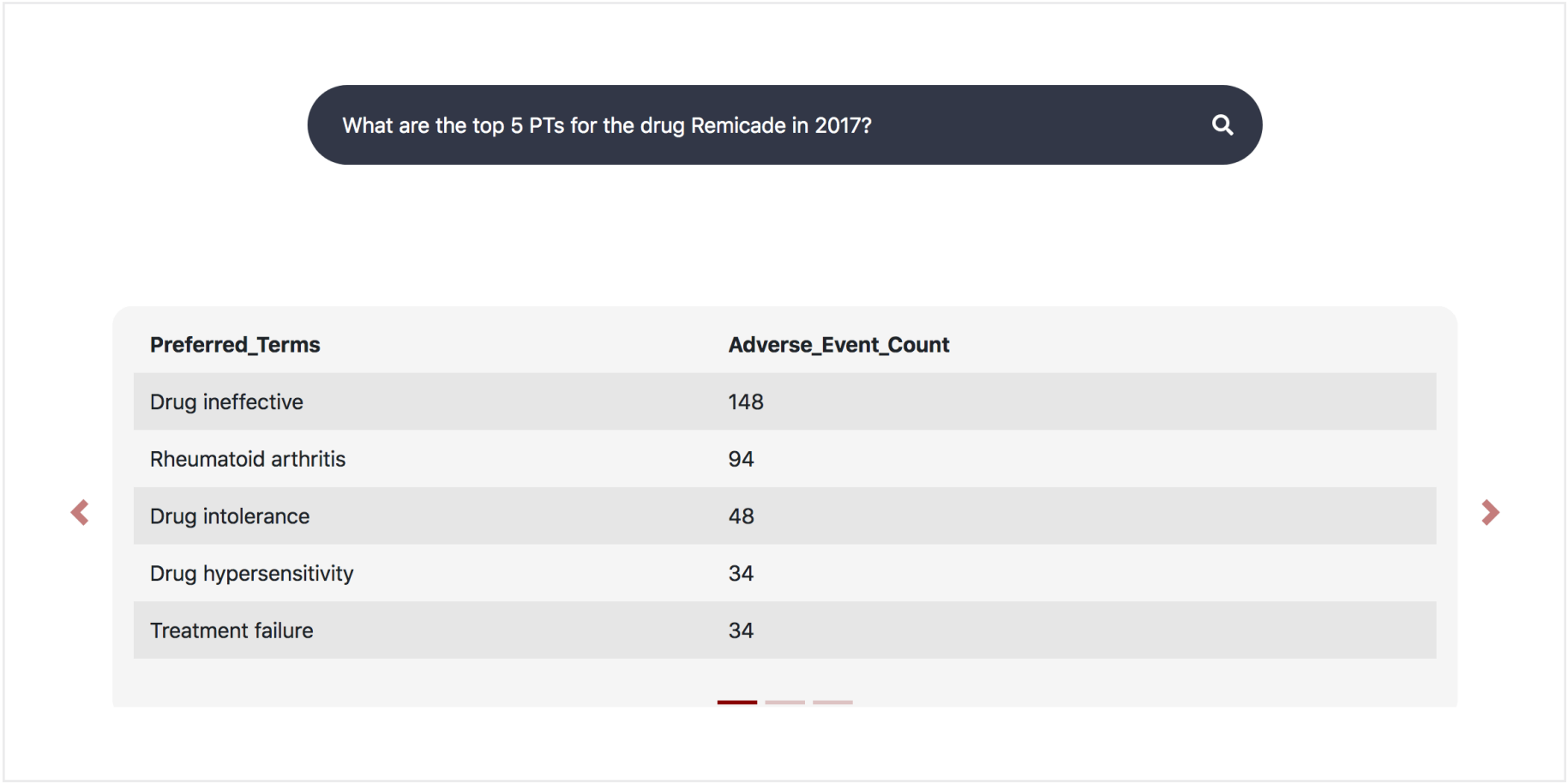
When conducting clinical trials, data is collected from electronic data capture systems and from central and local labs. This data is then combined and stored in a data warehouse. The data is transformed into a format in which the data managers are familiar. Data managers then review and clean the data. Once cleaned, the data is transformed into common data models (e.g., CDISC SDTM) and used for generating submission documents to the FDA. Machine learning models can be used to check patterns in data, and if there are irregularities or missing data, bring it to the attention of data managers for further review.

Data from prior studies is available for ML algorithms and models and can be learned from. Every clinical trial has certain milestones to achieve, and for each milestone, there are certain criteria to be met and associated documentation to be generated. ML models can be trained to understand if a trial is ready for a certain milestone. If not ready, it can determine the bottleneck, and it can make predictions for how long it will take to reach the milestone based on historical knowledge. It can generate the documentation needed for a milestone and send it to humans for approval before submitting to the FDA. For future clinical trials, the algorithms can answer questions such as, "How long will it take for me to enroll 'N' number of subjects for an oncology study?" and "How long will it take to reach a milestone based on the trial protocol document?"

Search

Data managers and reviewers log in to clinical data review platforms (CDRP), and slice and dice the data they want in order to review for missing, wrong, or inaccurate data. If they can search using natural language, they can spend more time on reviewing the data rather than creating complex search criteria. For example, users can write:

- "Show me all demography data where subjects are males, but pregnancy is yes."
- "Show me all data from adverse events and concomitant medications, highlighting concomitant medications without corresponding adverse events."
- "Show me all the severe adverse events reported for the third visit."



Preferred_Terms	Adverse_Event_Count
Drug ineffective	148
Rheumatoid arthritis	94
Drug intolerance	48
Drug hypersensitivity	34
Treatment failure	34

FIGURE 1: AN NLQ PROOF OF CONCEPT THAT PERFICIENT BUILT USING FDA AERS DATA

Confidence Scores

NLP/NLQ (natural language querying) converts these texts to search criteria, which is then converted to an SQL query. The query is executed, and the results are returned in figure 1.

When ML algorithms understand clinical data, they can execute the search criterion and deliver the results. Information extraction with language translation (e.g., English to ML) can be used. Extending this would include allowing audio input from users, converting audio to text, and then using NLQ to convert the text to corresponding queries. For non-English speaking users, NLP language translators can be used to achieve the same results.

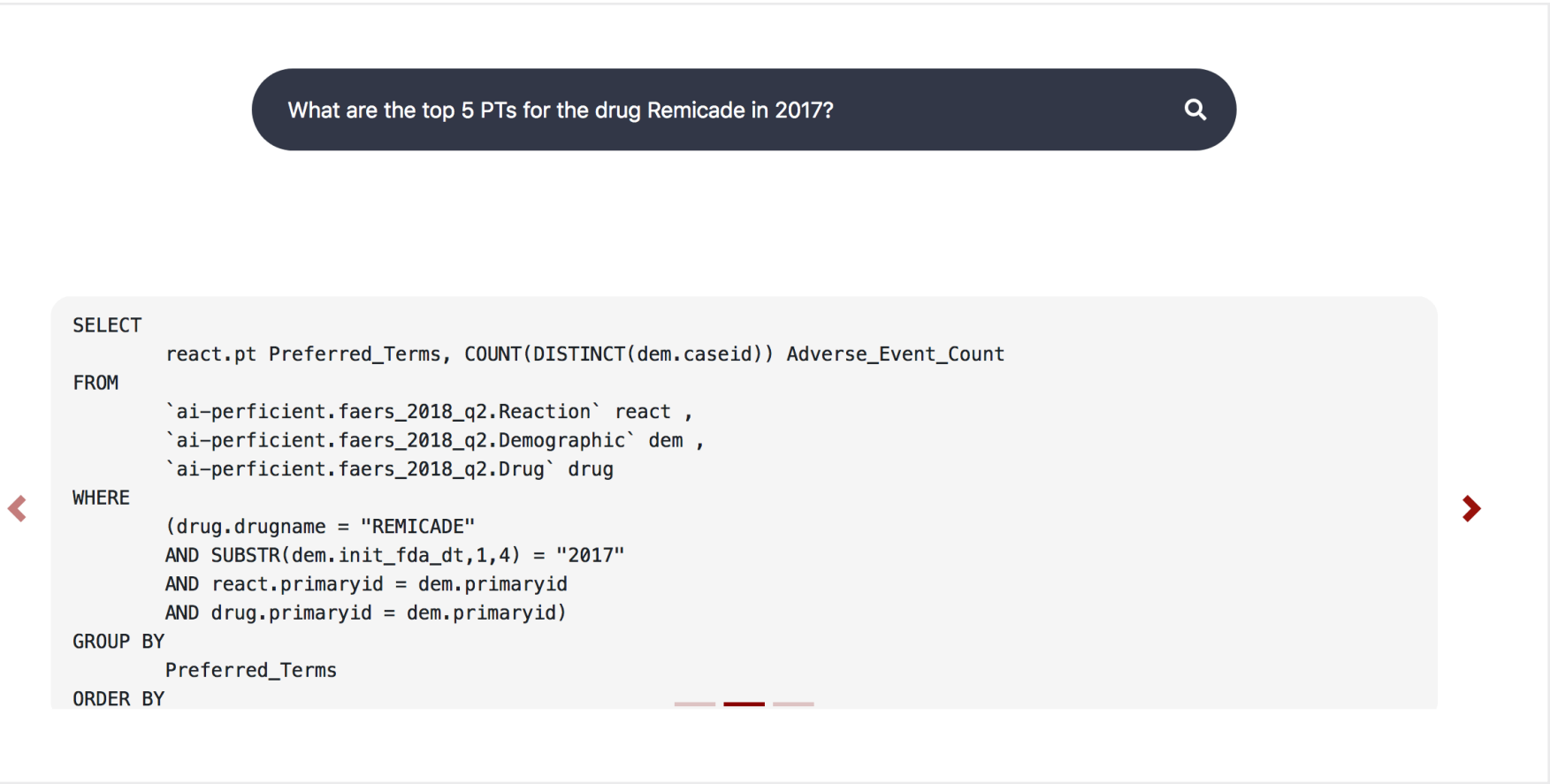


FIGURE 2: QUERY GENERATED USING NLQ

Data Review Prioritization

When clinical data reviewers, medical monitors, statisticians, and safety reviewers are reviewing data in the CDRP, an ML model can analyze the data and apply statistical analysis to determine the probability of whether the data is clean or not. In addition, this model can be used to detect anomalies in the data. Users should be presented with data based on the probability that the data is clean or the data points that require their attention. The data reviewers can prioritize their review activity based on the input from the ML models. It will be valuable for the users and will significantly reduce the time taken for data review. Users can set the threshold for the data points they want to review first, based on the statistical analysis done by the ML model.

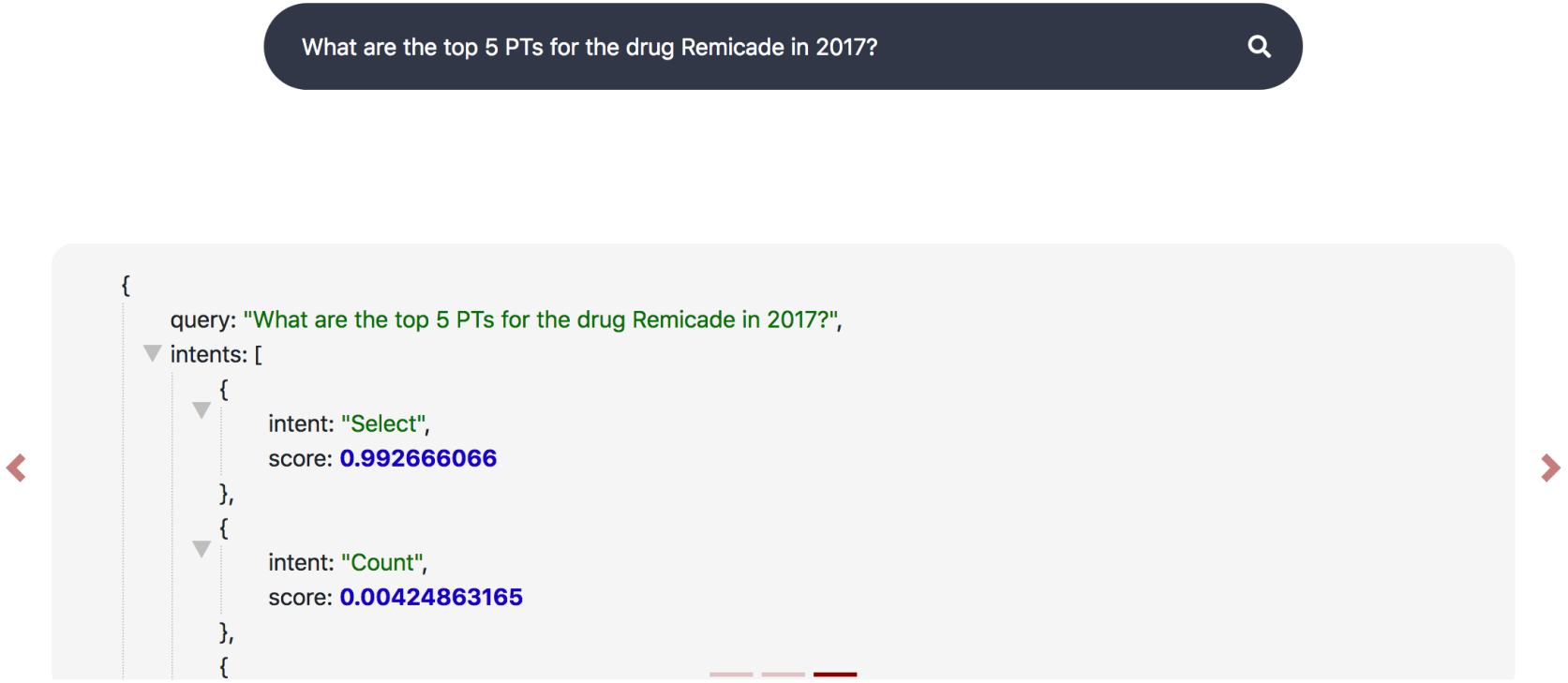


FIGURE 3: INTENT AND SCORE

Review Plans

Prior to starting data review, each team has its own review plan. The data review team has a data review plan, the safety team has a safety review plan, the analysis team has a statistical analysis plan, etc. These review plans are standard across studies with few changes. An automated program can create review plans based on the metadata information in a study. The review plan will result in a list of tasks, which can be assigned to different user groups. Based on the previously assigned tasks, the program can automatically assign the tasks and prioritize the tasks for each individual user. This, combined with the prioritized data review, will enable users to prioritize their tasks and complete the data review and the tasks from the review plans.

Statistical Analysis of Clinical Data

Statisticians analyze clinical trial data to understand the efficacy and safety of a drug. Machine learning can be used to assist statisticians in analyzing the data and looking for anomalies, such as:

- Why is it that the subjects in certain age ranges at site 1 are reporting fever in visit 3?
- Are there other sites that are reporting fever for visit 3?
- Is this related to the clinical trial, or is there an outbreak in site 1 region that is causing this anomaly?
- Can I compare this data with publicly available data to check for outbreaks?

There is enough data to train the ML algorithms to help in reviewing the data. Understanding how to use this data and how to perform more-effective exploratory analysis will help statisticians better understand the efficacy and safety of a clinical trial.

Robotic Process Automation

Validation is a manually intensive and time-consuming process. Some companies are considering using robotic process automation for validating their review platforms when new functionalities are added to it or when new versions of the software are released. This will ensure that the available functionalities in the current system are not impacted by new changes and no regression issues are introduced in new versions.

Success Criteria

Setting initial expectations and not promising a magic bullet is a key factor in determining the success of an initiative that focuses on deploying AI to streamline the clinical data review and cleaning process.

Training the machine learning model will determine how accurate the results are. After every phase, evaluating the released functionality, reassigning priorities to backlogged initiatives, and releasing based on prioritized functionalities should be done and closely monitored. The adoption by business users will determine how successful these initiatives are. Other factors that will help in evaluating how success include:

- Accuracy and speed of data review
- Effort needed by humans to reach a milestone
- Improved user experience
- Adoption of the solution by the end-user community

Extending the Roadmap

AI technologies can be used for more initiatives within CDRP. Each of the questions that need to be answered in the initiatives could be addressed by ML models based on how well the models are trained. In addition to CDRP, there are other areas in life sciences – such as safety systems, clinical operations systems, risk-based monitoring systems, and data capture systems – where AI technologies can be used to enrich the data management process.

About the Author

Prabha Ranganathan

Clinical Data Warehousing, Director, Perficient

Prabha works closely with life sciences companies to design and implement innovative clinical data review and cleaning solutions using a variety of technologies. With a clear understanding of the industry and strong technical knowledge, she brings a unique skill set to solve complex problems.

Prabha received her MBA from Babson College and an M.S, in Computer Science from Illinois Institute of Technology. She has also completed the Health Information Technology Leadership Development Program at Harvard and Professional Certificate Program in Machine Learning & Artificial Intelligence at MIT.

Let Perficient help you on your digital transformation journey.

Perficient is the leading digital transformation consulting firm serving Global 2000® and enterprise customers throughout North America. With unparalleled information technology, management consulting and creative capabilities, Perficient and its Perficient Digital agency deliver vision, execution and value with outstanding digital experience, business optimization and industry solutions.

PERFICIENT®
vision. execution. value.

PERFICIENT.COM/BLOGS

PERFICIENT.COM/GUIDES

(855) 411-PRFT(7738)

INFO@PERFICIENT.COM