

Single Day Event

Advancement and use of AI/ML within Pharma Industry

Sumit Kumar, *GenInvo*
Kshitij Vichare, *GenInvo*



Agenda

1. Introduction
2. About Data Science
3. Quality Checks for DocQc
4. E – protocol meta data extraction



Introduction

In the pharmaceutical industry and consumer healthcare business, AI and machine learning have played an important role. AI refers to the field of computer science that can analyze complex medical data. They are useful for detecting and diagnosing diseases, drug manufacturing, clinical trials, and predictive forecasting due to their ability to exploit meaningful relationships in data sets. Artificial intelligence provides the most significant advantage in reducing the costs related to drug development, in turn, enhancing returns on investment, and even contributing to an overall cost reduction for end users.



Accelerate drug discovery and development



Optimize and improve efficacy of clinical trials



Target specific patient's populations more effectively



Better insight into patient behavior to improve drug delivery and effectiveness and healthcare outcomes



Improve safety and risk management



Gain improved insight into marketing and sales performance



About Data Science

Data Science

Data science entails analyzing data to find patterns and characteristics. A combination of mathematics, statistics, and computer science are involved here. A variety of visualization techniques are used to make the data understandable. The focus here is to gain insights that can make a positive difference to the organization by understanding and using the data.

Artificial Intelligence

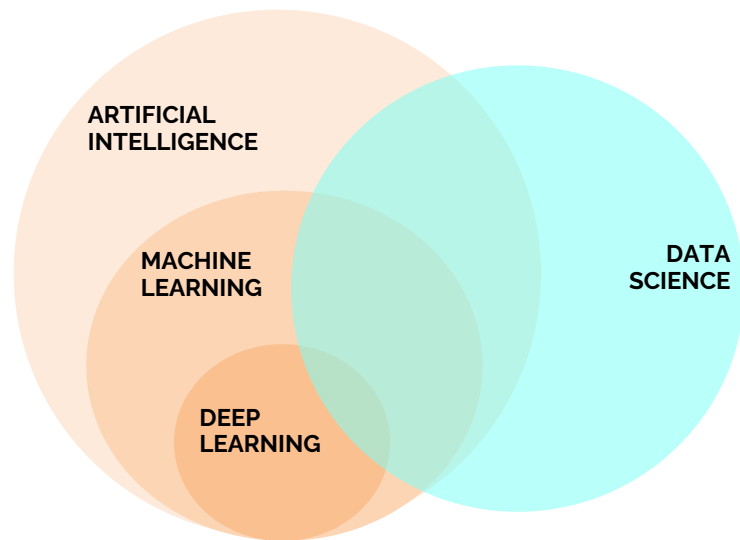
Like natural intelligence, artificial intelligence enables computers to mimic the human intelligence and behavior.

Machine Learning

Machine Learning is a branch of Artificial Intelligence that focuses on learning. In order to learn autonomously from data, algorithms and statistical models are developed without explicit instructions being given up front.

Deep Learning

Deep Learning is part of Machine Learning that uses artificial neural networks. Models of this type are based on the structure and function of the human brain.





CSR QC Manual Process

Existing Process & Challenges

A major reason for delayed trial data reporting is CSR-QC. QC is currently done manually, which is inadequate and presents significant challenges in terms of time, quality, and management of resources.

- An extensive manual quality check was involved in the complex study report involving extensive numerical data.
- There is a high likelihood of observational errors during the CSR review cycle, which delays the overall turnaround time for producing documents

Most critical and time-consuming checkpoint for CSR-QC is verification of numerical number taken from source tables. Medical writers manually create several in-text tables in CSRs. There are several sources of Tables, Figures and Listings (TFL) generated by the biostatistician that need to be compared to these in-text tables. CSR-QC is the most important part of clinical trials, as it has a direct impact on clinical trial outcomes as well as patient safety. This is currently done manually, which not only takes a lot of time but also causes observation errors.



DocQC Intelligent Automation

Developing a concept

DocQC was conceptualized in light of the complex and repetitive process of verifying source data against supported data (Protocol, Template, Source Tables, References, etc.), which is prone to observational error and data discrepancies.

Several CSRs and supporting documents were reviewed in an effort to understand the possible data extraction pattern for in-text and source tables. By using a rule-based, machine-learning driven cognitive search algorithm, data can be retrieved from the CSR, mapped into the right section, then compared to present the differences. In the overall quality check process, an automated solution could potentially save time and effort. Hence, DocQC was born.

Challenges

- Extraction of Data
- Data Mapping with supported Documents
- Reading complex RTF tables
- Grammatical Issues

Solution developed for Checks

- Abbreviation checks
- Citations & Post text TFL
- References checks
- SAP Protocol check



ApoGI e-Protocol Module

Auto extraction of Meta data from Protocol documents

1. Meta data such as Trial design datasets, study visit schedule etc., forms the basis of CRF generation and SDTM data transformation. At present ,this meta data is being extracted manually from the protocol.
2. This module aims to auto extract the meta data using NLP techniques such as NER model, Q&A model and semantic search.



Text Extraction

- Pdfminer
- Grobid



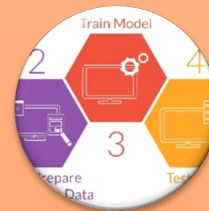
Table Extraction

- Pdf plumber
- Tabula



Information Retrieval

- NLP



Preparation of training data

- Training data been collected from various publication





The Global Healthcare Data Science Community

Contact Channels

- 📞 UK +44 1843 609600
- 📞 USA +1 609 514 5105
- ✉️ office@phuse.global
- 🌐 phuse.global

Social Media

- 🐦 [@phusetwitta](https://twitter.com/phusetwitta)
- 📘 [/phusebook](https://facebook.com/phusebook)
- ▶️ [/phusetube](https://youtube.com/phusetube)
- 🌐 [/company/phuse](https://linkedin.com/company/phuse)