DAY 2: 03-DEC-2022
Parallel Track – 4 (2.00 PM to 3.30 PM IST)

Unveiling the impact of AI initiatives in Clinical Data Management
Session Chair and Speakers

Santosh Karthikeyan
Session Chair

Kedar Deshpande
Speaker

Ayush Mittal
Speaker

Soumya Veerla
Speaker

Abhishek Kadam
Speaker

Shreyans Patel
Speaker
<table>
<thead>
<tr>
<th>S. No.</th>
<th>Agenda</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Slido – Have your say</td>
<td>15 mins</td>
</tr>
<tr>
<td>2</td>
<td>AI influencing Quality and Efficiency – Kedar &amp; Ayush</td>
<td>20 mins</td>
</tr>
<tr>
<td>3</td>
<td>AI Assisted Data Review for Significant Efficiency &amp; Accuracy - Soumya</td>
<td>20 mins</td>
</tr>
<tr>
<td>4</td>
<td>Under the Hood, AI Empowers CDM – Abhishek &amp; Shreyans</td>
<td>20 mins</td>
</tr>
<tr>
<td>5</td>
<td>Q&amp;A</td>
<td>10 mins</td>
</tr>
</tbody>
</table>
Have your say

Please open https://www.slido.com in your web browser and provide the code #SCDMINDIA or scan the below QR code.
Presentation (Kedar & Ayush)
Artificial Intelligence (AI) can enhance clinical trials and augment human efforts by enabling quick and easy access to data and documents. AI powered tools and solutions can fully adapt to user needs and can automate, accelerate, and optimize processes in clinical trials.

The availability of an AI-enabled tool that allows users to upload, select, search, compare, and download required information from clinical trial documents can bring much-needed change and cause a revolution in the clinical research field.
Problem Statement

Pharma, CRO, BPO and Tech companies are required to generate number of delivery ready artifacts that requires vetting by multiple stakeholders.

Historical Data Store
- Absence of organized and centralized historical data.
- Reference for a new project with similar endpoints and success factors.
- Need analytical insights to drive decision making

Time Investment
- Large number of regulatory compliant, submission ready documents.
- High Expense. Profitability impact.

Dependency
- Readiness of upstream documents and processes on downstream activities.
- Manual approach to creation of EDC objects like CRFs and edit checks.

Review/Approval
- Systematic approach to review and approval.
- Accelerate finalization of artifacts.
AI Framework for Clinical Trials domain

- **NLP enabled** protocol content ingestion
- **AI/ML powered** tool that learns and matures.
- **Knowledge store** integrated with external data sources like CTGov.

- **Ontology** driven knowledge graph
  - Locate projects with identical end points and success factors.
  - Analytical data driven decisions.

- **Auto-mapping** source to target structures
  - Systematic generation of modelling scripts

- Automated creation of ready to use content for **targeted individuals/groups** e.g. Medical writers, Bios, CRF designers.

- **User Friendly Interface/ Enhanced User Experience**
  - Easy to use, responsive.
  - Create objects within seconds.
  - GxP Compliant validated system

- Content Management with workflows
- Content creation and content prediction with **auto text generation models using NLG**
- Paraphrasing, text suggestion, text completion
Proposed Framework

Data Sources
- CTGov
- Structured Data Sources
- Documents
- Video/Images

Workflows
- Role based access
- Approvals
- Notifications
- Collaborations

Machine Learning Layer
- Document Parser
- Ontology Builder
- Text Generation
- Semantic Mapping
- Active learning

Ontology Builder
- Encoder-Decoder
- Graph Neural Network
- Transformers
- Transformers
- Transformers

Process Flow
- Upload Protocol
- ML/NLP/NLG
- Extract Contents
- Database Storage
- Ontology
- Knowledge Graph

Knowledge Store
- Insights
- Exclusion – Inclusion Criteria

Granularity

Documents

eCRF Generation
- ✓ eCRF design
- ✓ EDC upload able file

Edit Check originator
- ✓ Edit Check Specification Creation
- ✓ Attach Edit Check Program

SDTM
- ✓ SDTM Annotation
- ✓ SDTM Mapping Specification
- ✓ SAS Program

Dummy Data
- ✓ Dummy EDC dataset

Test Case Generation
- ✓ Test Case Generation for edit check testing.
- ✓ Test Case Automation

Analytics
- ✓ Protocol Analytics
- ✓ Site Intelligence
- ✓ Statistical Analysis

Documents
- ✓ Regulatory Documents
- ✓ CSRs/SAPs/Protocols
- ✓ Manuscripts/Publications
Knowledge Store

- Exhaustive Knowledge Store is Key.
- Some examples we captured in this framework:
  - Documents
  - ECRFs
  - Insights
  - Medical Codes
  - Objectives
  - Endpoints
  - Time & Event Schedule
  - CTGov Data
  - CDISC
  - Exclusion Inclusion Criteria
ML Modelling

- **Document-Net**
  - Self-supervised pre-training
  - Masked Language Models
  - Transformers (Multi-modal)
  - Word Embeddings + Image Representation

- **Building Knowledge Store**
  - Graph Neural Network
  - Matching Networks
  - Encoder-Decoder Architectures
  - Semantic Mappers

- **Text & Analytics Generation**
  - Transformers (Multi-modal)
  - Wide and Deep Neural Networks
  - Diffusion Language Models

- **Active Learning**
  - Embedding Stores
  - Intelligent Semantic Cache
  - Feedback driven explore-exploit network
Leap Forward

1. Study Protocol
   - Week 1: Protocol Creation
   - Week 2: Protocol – Creation | Review | Finalization
   - Week 3: Protocol Finalization
   - Traditional Timelines
   - Proposed Timelines

2. Study Build
   - Week 1: eCRF Design
   - Week 2: eCRF Build | eCRF Finalization | Edit Check Spec Creation | Test Case Programming
   - Week 3: Edit Check Spec Finalization | Starting Edit Check Program | Validation | Go Live / FPI
   - Week 4: Test Case Generation
   - Traditional Timelines
   - Proposed Timelines

3. SDTM
   - Week 1: SDTM Mapping Spec V0.1
   - Week 2: Review
   - Week 3: SDTM Mapping Spec V1.0 | Starting SDTM Program
   - Week 4: SDTM Program V0.1
   - Week 5: SDTM Program V1.0
   - Traditional Timelines
   - Proposed Timelines

Weeks Saving:
- eCRF: 2 weeks
- Study Build: 8 weeks
- SDTM: 10 weeks

Conclusion

• As competition in the market for products and services intensifies, companies are looking to automation as a solution in their quest to reduce operational costs.

• A lot of manual effort can be saved by digitally accessing the clinical study documents like protocols, supported by NL queries, NLG, ML and Knowledge Graphs.

• The combination of AI-powered tools and the wonder of the human brain has the potential to revolutionize the way documents, analytics and other artifacts are generated today in the clinical research field.

• Proposed framework can significantly reduce time required for generating high quality deliverables.
Presentation (Soumya)
AI Assisted Data Review for Significant Efficiency & Accuracy

Soumya Veerla
Senior Manager, Data Management
ICON PLC
Aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities.

GPAI brings together engaged minds and expertise from science, industry, civil society, governments, international organizations and academia to foster international cooperation.

29 countries
A computer program/system that does something we would normally think of as intelligent in humans.

AI technologies extract concepts and relationships from data and learn independently from data patterns (clinical data), augmenting what humans can do (reconcile, identify discrepancy, raise query) and interacting with humans in a natural way.

- Machine learning,
- Deep learning,
- Un/supervised learning,
- Natural language processing (NLP)
Journey – AI in Data Review

4. Solution?
Idea was to innovate & create inhouse scalable product

5. Solution Finalized?
AI programme with Python, machine learning, combination of NLP & rules-based pattern matching

6. Proof Of Concept (POC)
CM-AE-MH Review
3 studies

1. Can we automate?
Manual review of clinical data listings which need critical thinking and human cognitive decision making

2. Which listings to choose?
High volume & Impact
Maximum time & effort
Generic- across all studies, TA, Phase

3. Explore Solution?
Did not go for off the shelf product available in market
• Artificial Intelligence, mainly NLP developed to review Con Med, Medical History & Adverse Event.
• Combination of NLP & rules-based pattern matches indication from CON Med to Adverse event or medical history along with corresponding dates.
• Continuous update of programme from user feedback

• AI assisted in identifying each of the records which needs action or no action as per match.
• AI output Action required and no action required was reviewed 100% manually to check accuracy and efficiency of program.
• After each round of review feedback was provided to development team.
• Program started learning and correcting itself giving higher percentage of accuracy with every new round of review.
AI output:
Good Match – No review required: When there is no discrepancy
No Match DM review required: When there is discrepancy in date or term
<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Medication Name</th>
<th>CM Start</th>
<th>CM End</th>
<th>Indication</th>
<th>AE Term</th>
<th>AE Start</th>
<th>AE End</th>
<th>AI Output</th>
<th>AI Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXX</td>
<td>CALCITONIN</td>
<td>7-Mar-22</td>
<td>7-Mar-22</td>
<td>HYPERCALCEMIA</td>
<td>HYPERCALCEMIA</td>
<td>7-Mar-22</td>
<td>9-Mar-22</td>
<td>Good Match</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>VOMEX</td>
<td>14-May-21</td>
<td>19-May-21</td>
<td>NAUSEA</td>
<td>NAUSEA</td>
<td>13-May-21</td>
<td>17-May-21</td>
<td>Good Match</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>PARACETAMOLE</td>
<td>29-Oct-21</td>
<td>29-Oct-21</td>
<td>Flu</td>
<td>Flu</td>
<td>30-Oct-21</td>
<td>4-Nov-21</td>
<td>No Match</td>
<td>Date Mismatch</td>
</tr>
<tr>
<td>XXXX</td>
<td>FENTANYL</td>
<td>27-Aug-22</td>
<td>27-Aug-22</td>
<td>ANESTHESIA FOR PCN REPLACEMENT</td>
<td>RENAL AND URINARY DISORDERS- OTHER; PCN DISLODGED</td>
<td>27-Aug-22</td>
<td>30-Aug-22</td>
<td>No Match</td>
<td>Term Mismatch</td>
</tr>
<tr>
<td>XXXX</td>
<td>AMOXICILLINE + CLAVULANIC ACID</td>
<td>21-Dec-21</td>
<td></td>
<td>feverlachycardia</td>
<td>Fever</td>
<td>21-Dec-21</td>
<td></td>
<td>No Match</td>
<td>Multiple Terms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Medication Name</th>
<th>CM Start</th>
<th>CM End</th>
<th>Indication</th>
<th>MH Term</th>
<th>MH Start</th>
<th>MH End</th>
<th>AI Output</th>
<th>AI Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXX</td>
<td>LORAZEPAM</td>
<td>14-Jul-20</td>
<td></td>
<td>ANXIETY</td>
<td>ANXIETY</td>
<td>18-Jun-20</td>
<td></td>
<td>Good Match</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>ENDONE</td>
<td>UN UNK 2020</td>
<td></td>
<td>arthritis</td>
<td>ARTHRITIS</td>
<td>UN UNK 2020</td>
<td></td>
<td>Good Match</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>OZEMPIC</td>
<td>UN Jan-21</td>
<td>1-Nov-21</td>
<td>diabetes</td>
<td>DIABETES</td>
<td>UN UNK 2015</td>
<td></td>
<td>Good Match</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>RABEPRAZOLE</td>
<td>21-Jul-19</td>
<td></td>
<td>GERD</td>
<td>HYPERTENSION</td>
<td>UN UNK 2015</td>
<td></td>
<td>No Match</td>
<td>Term Mismatch</td>
</tr>
<tr>
<td>XXXX</td>
<td>BISOPROLOL</td>
<td>UN-Aug-18</td>
<td>11-Jul-22</td>
<td>ischemic Heart Disease peripheral Artery Disease</td>
<td>ischemic Heart Disease</td>
<td>UN-Aug-18</td>
<td></td>
<td>No Match</td>
<td>Multiple Terms</td>
</tr>
<tr>
<td>XXXX</td>
<td>LEVETIRACETAM</td>
<td>25-Sep-22</td>
<td></td>
<td>aphasia</td>
<td>APHASIA</td>
<td>7-Apr-22</td>
<td>7-Apr-22</td>
<td>No Match</td>
<td>Date Mismatch</td>
</tr>
</tbody>
</table>
POC Outcome

**Manual Review**

100% Effort – 0% Efficiency  
High turn around

**AI Assisted Review**

15% Effort - 85 % Efficiency  
Quick turn around

![Bar chart showing comparison between Manual Review and AI Assisted Review](chart.png)
NLP has the ability interpreting multiple data points – symptoms, medications, start & stop dates from various listings and transforms it into algorithm to identify actions required in case of discrepancy and no action required when data is reconciled, similar to critical thinking and human cognitive decision making. ML after each review updates algorithm to make it robust.

We are persuaded with the outcome of POC and the project is in the implementation phase. The program developed can be easily implemented across all studies.
Presentation (Abhishek & Shreyans)
Under the hood, AI empowers CDM !!!

DISCLAIMER: All opinions expressed in this presentation and on the following slides are solely those of the presenter, and do not reflect the views or opinions of Novartis. Neither SCDM will not be held liable for the content presented. All the images in this presentation are taken from internet. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Presenter:
Abhishek Kadam, Associate Director, Novartis
Shreyans Patel, Associate Director, Novartis
Unveiling the Case of Nora and John

Nora is a Data Manager and has a deeper experience of implementing AI/ML solutions in her studies for data management activities.

John, study data manager is new to AI/ML concepts and wants to understand how these concepts can impact his way of working.

Let’s see what John learns from Nora about Impact of AI/ML on Clinical Data Management
Let me explain you where the industry is using these technologies in Clinical Data Management......
Current Application of AI in CDM

- **Identify KRIs, KPIs**
- **CTTI – CTQs**
- **LSTM, SVM**
- **Predictive KRIss & KPIs For Milestones**
- **Big Data Reviews**
- **Risk Data Review Insights, Trends & Patterns**
- **Audit Trails, Real World Evidence, eSource data**
- **Isolation Forest**
- **Anomalies, Data Insights, Trends, Patterns, Outliers**

**Data Visualization**
- Tools, Integrations
- Dashboards, System Alerts

**Setting Up Enterprise Tools**
- **Database Elements Recommender**
- **Extract Protocol Information**
- **Apriori Algorithm**
- **Time & Event Structure, Similarity Algorithms**

**Before FPFV**
- **Risk Based Database Set-up**
- **Literature & Specifications Mining + Virtual Site Assistant for EDC**

**After FPFV**
- **AI Based Query Generation & Closure**
- **Key Decision Makers**
- **Identify issue, write query Post Query**

**CRFs, Fields, Edit Checks, Manual Checks**

**SVM – Support Vector Machine, LSTM – Long Short Term Memory**
Let’s go through how AI demands better infrastructure and technology....
Impact on Infrastructure & Technology

Training NLP Model in Regular Laptop vs on Cloud Sever with required GPUs

8 hrs. for approx. 3 million records

5 min. for approx. 3 million records

NLP – Natural Language Processing
Impact on Infrastructure & Technology

**Ecosystem (01)**
Integration of AI tools with DM tools
Technology stack with AI/ML and Data science work – TensorFlow, Keras etc.
Code maintenance through Git or similar platforms
Prototyping environments and sandboxes, IDEs like PyCharm, R-Storm, Streamlit, etc.

**Compute Engines (02)**
- GPUs, Clouds
  - For scalable solutions
  - Distributed file system and RDDs to work on big data
  - GPUs for high compute power and processing of big data

**Platforms (03)**
Cloud service providers
- Leveraging the state-of-the-art platforms created for AI ML development
- Platforms enabling ML Ops and Low code environments

**Tools (04)**
Technological, Integrated systems
- Commercial off the shelf solutions can be utilized for programming and visualization

One stop shop for all these infrastructure needs is a emerging impact of AI ML on clinical data management infrastructure.
After this infrastructure, in my study, Can I merge X-Ray Images, MRI recordings and sensor data with CRF datasets for Data Review by AI?

Sure! Consider these Data Engineering and Data Privacy aspects. Close collaboration between these team is a key!!
Now let us see how Clinical Data Management skill sets are influenced by application of AI in CDM!
Skill Development – AI & Data Science

A Mix of Life Science + Data Science Experts will be required in DM!!

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Details</th>
</tr>
</thead>
</table>
| New Skill Sets/Reskilling in Data Management | Programming – Python, R etc.  
Tools Knowledge – Analytics tools and Low code tools  
IDEs – PyCharm, VS Code etc. |
| Training Curriculum                   | Coding with Python or R  
Basic Data Science Concepts – Supervised vs Unsupervised, Deep Learning  
Interpreting Visualization – Histograms, Scatter Plots |
| Recruitment Strategy                  | Team - A mix of Data Science and DM Experts                           |
| New Roles & Career Development        | Evolution of Clinical Data Scientist, Data Analyst, Data Translator, Data Engineers in CDM |
Collaborating with partners

FSP Partners
Provide resources skilled in Data Science, AI and AI enabled tools

Recruitment Partners
Recruiting & retaining niche talent in shorter duration of time

Technology Partners
Niche partners who have capabilities to support AI/ML Journey.

Training Partners
Specialized trainers for delivering Niche content in AI/ML Data Science Trainings in simplified /customized manner
Impact on Regulations & Processes

Finally, AI has an impact on the CDM process and regulations. Let us see how!
Impact on Regulations and Processes

**Impact on Regulations**
- New FDA regulations for validating SaMDs
- EU & other regional Data Privacy Guidelines

**Functional Impact**
- Creation of New Process Maps
- Getting away from 100% Data Cleaning
- Additional Data Scrutiny on Audit Trials, RWE data etc.

**Impact on Decisions**
- Responsibility with Machines & Accountability with Humans
- Reliance on Algorithm outputs for decision making

- Use of differential privacy by technology giants
- Emergence of the AI enabled query management
- Human in the loop systems assisting decisions
“AI is probably the most important thing humanity has ever worked on. I think of it as something more profound than electricity or fire.” – Sundar Pichai.
References

• The automation of CDM Driven activities (Version #1), SCDM Innovation Committee –CDS Topic Brief

• Utilizing Artificial Intelligence for Efficient CRF design (lexjansen.com)

• https://www.ctti-clinicaltrials.org/projects/quality-design


• Artificial Intelligence (AI) in Clinical Trials Market is (globenewswire.com)
Q&A
Thank You