Paper ML18

The Best of Both Worlds: Machine Learning and Human-in-the-Loop Processing in TLF Design, Production, and Validation

Ilan Carmeli, Beaconcure, Boston, Massachusetts, USA
Steve Ross, Beaconcure, Raleigh, North Carolina, USA

ABSTRACT

Despite misguided proclamations that (eventually) AI will render humans irrelevant in the clinical research analysis and reporting process, a combination of automation and human-in-the-loop (HITL) oversight in the creation and maintenance of a robust metadata layer between ADaM and TLF remains the single best way to enhance the accuracy and reliability of AI systems. In this paper we examine the integration of machine learning techniques such as Generative AI combined with Human-in-the-Loop (HITL) processing, specifically in the design, generation, and validation of clinical tables. By recalibrating the roles of man and machine in TLF development, statisticians, programmers, and data scientists can ensure the highest quality specifications and metadata while benefiting from the automation and scalability of AI techniques in drug development. We will explore production applications, demonstrating how this approach leads to more accurate outputs; faster, more context-rich validation, and optimized metadata flows. Attendees will gain insights into streamlining their workflows through the balance of human expertise and AI technologies.

INTRODUCTION

The increasing complexity and scope of modern clinical trials demand that we continually revisit and refine how we create and validate statistical displays. The analysis and review of clinical trial data has traditionally been a time intensive, manual process. Growing adoption of new technologies such as Generative AI and LLM (Large Language Models) have introduced new avenues for shifting the man-machine balance in task assignment to optimize workflows.

The current challenges in clinical trial analysis stem from several interconnected factors that contribute to inefficiency. Increasingly complex trials and “big data” strain our ability to produce and review outputs in a shorter timeframe. Rising regulatory requirements add deliverables to our workload. The normal involvement of multiple stakeholders across an organization creates intricate communication channels which are, by design, broken into silos, blunting their utility and effectiveness. Similarly, an iterative review process creates bottlenecks in the drug development pipeline.

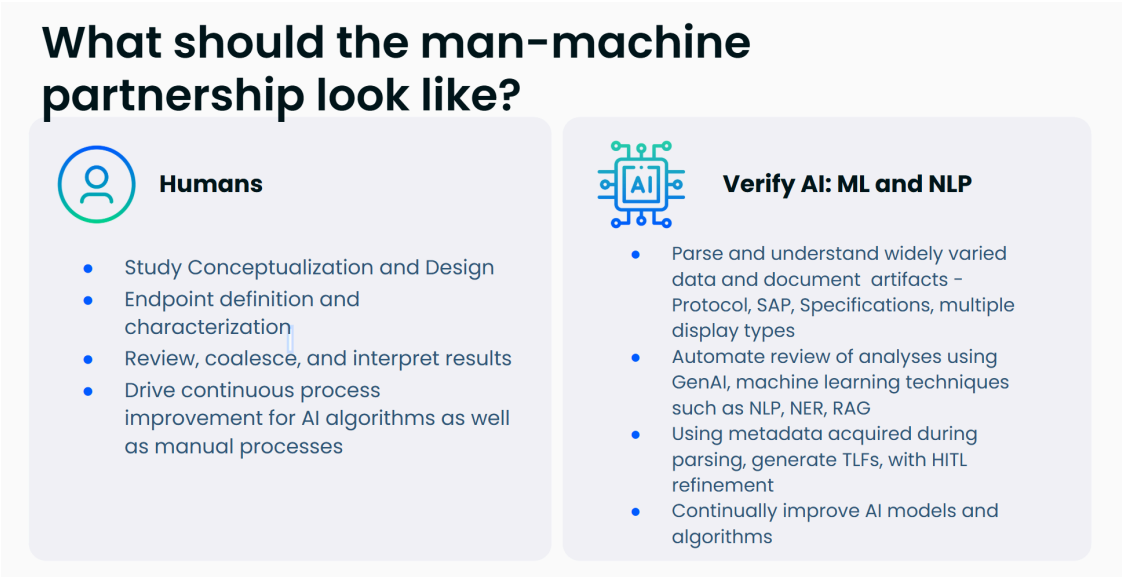
The way forward combines process reengineering with strategic AI implementation. Rather than treating AI as a complete solution, we propose a dual approach that modernizes workflows while focusing AI's capabilities where they add the most value. The integration of process innovation and targeted AI implementation creates a multiplier benefit that neither approach would achieve independently. This balanced approach ensures that technology serves as an enabler of human expertise rather than attempting to replace it, resulting in more efficient and effective statistical analysis processes.

THE INTEGRATION OF AI AND HUMANS-IN-THE LOOP

In today’s TLF development scheme, entrenched processes continue as if technology has not changed in decades. Programmers independently program; reviewers follow up with layers upon layers of sequential and iterative review. Subject matter experts - MDs, PhDs, PharmD’s who are paid to understand statistics, complex medical processes, pharmacokinetics and drug safety, or scientific writers who are experts in telling the ‘truth behind the numbers’ iterate through TLF sets, first programmatically, then manually to locate and iron out inconsistencies, questionable calculations, or bad data. Along the way, they note formatting issues, perform logic checks for things like horizontal and vertical summations, cross-table checks to make sure like metadata have the same numbers, et al. Shouldn’t your most

valuable minds focus on the *interpretations* rather than the things that can be mastered by the "machine" part of your team? Of course, the answer is yes. In some cases the solution to a problem is technological, other times it is simply procedural. For TLF production, the solution is both.

THE MAN-MACHINE PARTNERSHIP

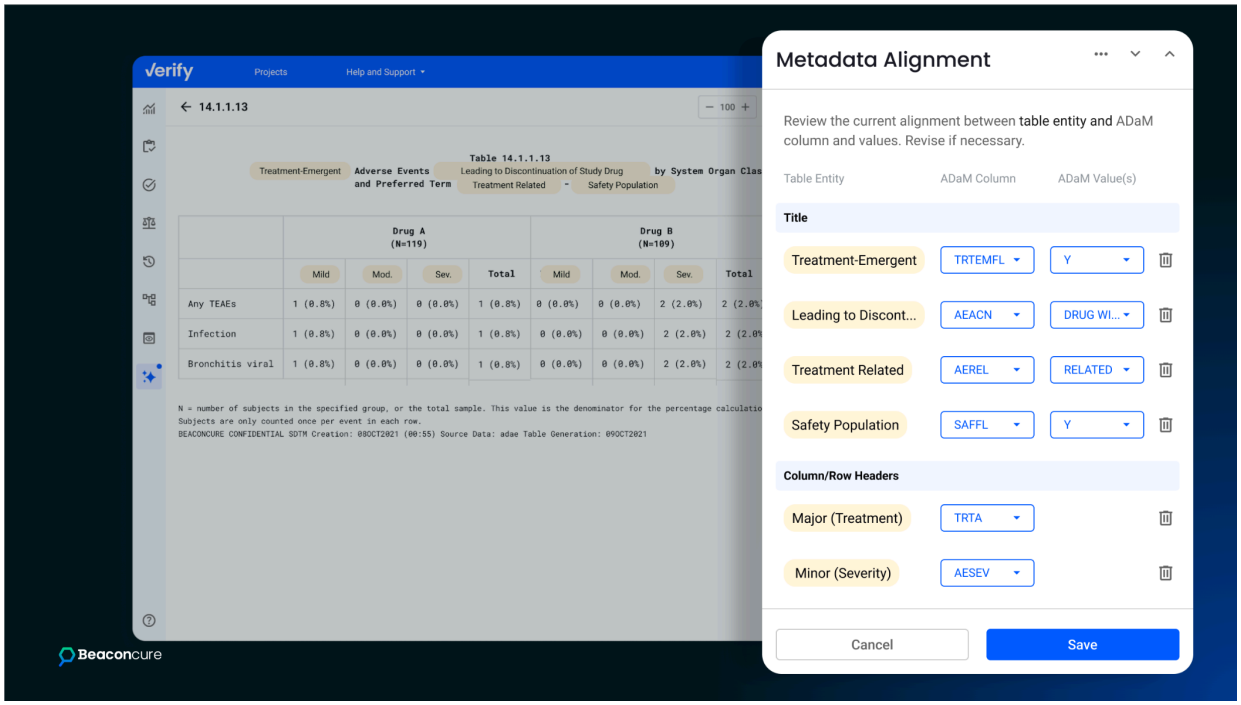


The optimal division of responsibilities between human experts and AI systems has changed as computational power has grown. Certain tasks are better suited to resolution by human expertise, while others can be effectively automated through AI systems.

Human expertise remains irreplaceable in several areas. Study conceptualization and design require deep understanding of clinical implications and research methodology that only experienced professionals can provide. The definition and characterization of endpoints demand nuanced interpretation of clinical significance and practical feasibility. Humans excel at reviewing, coalescing, and interpreting results in context, drawing upon years of experience and domain knowledge. Furthermore, they play a vital role in driving continuous process improvement for both AI algorithms and manual processes, ensuring that systems evolve to meet changing needs and requirements.

The AI-enabled functions complement human expertise by handling tasks that benefit from automation and consistent, rapid processing. Advanced systems excel at document digitization and parsing, processing large volumes of data with high accuracy. Entity extraction and relationship mapping can be performed systematically, creating comprehensive networks of interconnected information. Automated validation checks ensure consistent application of rules and standards, while discrepancy identification helps focus human attention where it's most needed.

ADDING VALUE THROUGH THE METADATA CURATOR

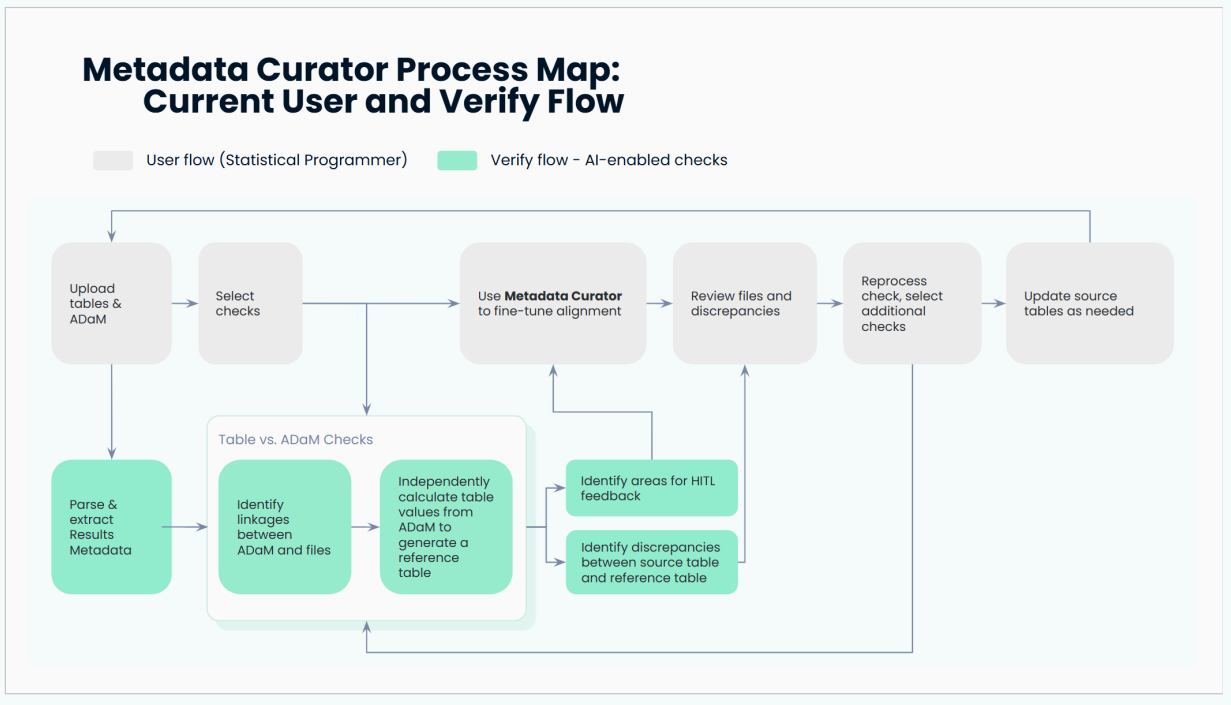


The insertion of a metadata curator step between the machine learning discovery/parsing process and TLF generation allows for the “Humans-in-the-loop”(HITL) to act as the guarantor of the machine in the process, ensuring the accuracy of the metadata. This step incorporates several key components that work in concert to ensure optimal outcomes.

The user flow for statistical programmers begins with the upload of tables, specifications, a Table of Contents, and ADaM datasets into Verify, followed by careful selection of appropriate validation checks. A combination of techniques including Named-Entity-Recognition(NER), Retrieval-Augmented-Generation(RAG) and rule-based methods combine to extract available metadata from TLF sets, specifications, and tables of contents in order to create a robust metadata layer that can be used for both current and future study development. Programmers as HITL engage in fine-tuning metadata through the curator interface, reviewing discrepancies as they arise and refining outputs based on their expertise. This human-directed process ensures that all automated results align with analysis plans, CDISC standards, and good programming practice.

AI-enabled verification works alongside human oversight, providing comprehensive support through semantic understanding of table structures and summary data. The system handles parsing and extraction of results metadata, identifying linkages between output files and other ‘referent’ documentation. It independently calculates reference tables for validation purposes and automatically detects discrepancies that require human attention. This automated support structure allows human experts to focus their attention on critical decision-making rather than routine verification tasks.

EVOLUTION OF CLINICAL RESEARCH ANALYTICS



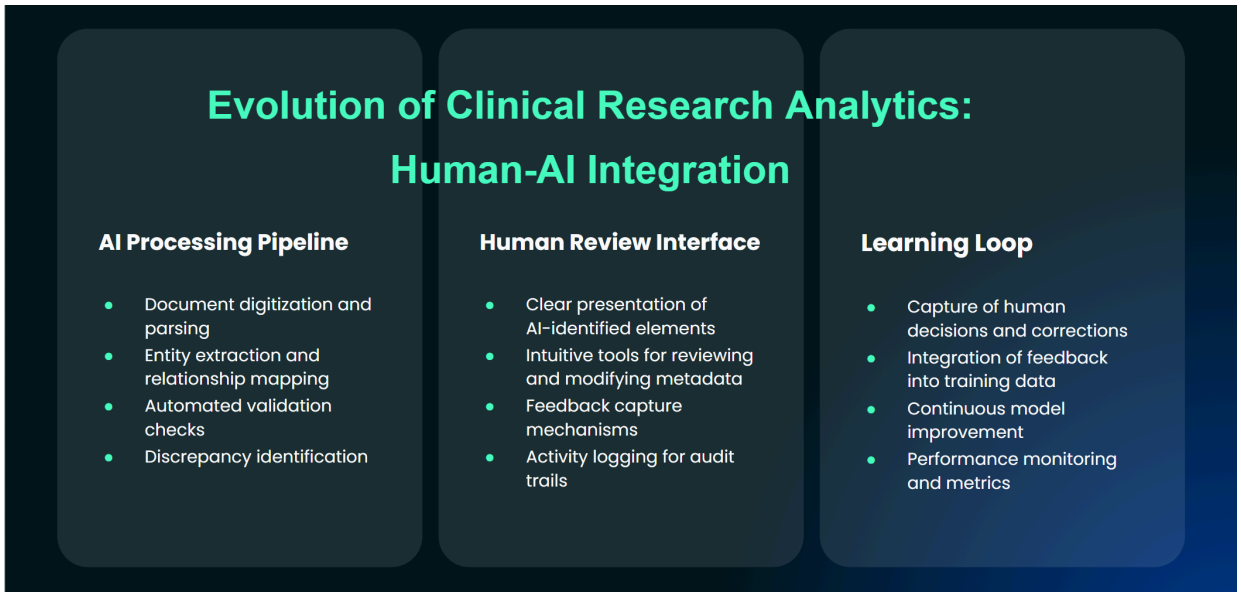
This implementation of clinical research analytics represents a sophisticated three-pillar approach to modern trial analysis, each component working in harmony with the others to provide a robust, accurate, and repeatable output and metadata package.

First, the **AI processing pipeline** forms the foundation of automated data handling, including comprehensive document processing capabilities. This system manages document digitization and parsing with high efficiency, while performing sophisticated entity extraction and relationship mapping between outputs and reference documentation such as tables of contents and TLF specifications. Verify functionality also includes robust automated validation checks and discrepancy identification processes, ensuring consistent and reliable initial processing of statistical displays.

Next, the **human review interface** serves as the bridge between automated processes and HITL oversight. This interface presents AI-identified elements and provides intuitive tools for reviewing and modifying metadata as needed. The system captures feedback systematically and maintains comprehensive activity logs for audit purposes, ensuring full transparency and traceability in all decisions.

Finally, the **learning loop** represents the system's capacity for continuous improvement and adaptation. This mechanism captures human decisions and corrections, integrating this feedback into the training data that powers the AI model. Through continuous model improvement and performance monitoring, the system becomes increasingly accurate and efficient over time, while maintaining the flexibility to adapt to new table types and document structures.

BENEFITS AND OUTCOME



The implementation of this integrated approach yields measurable benefits across three key dimensions, each contributing to improved efficiency and accuracy in clinical trial analysis.

Speed improvements manifest in several crucial areas of the analysis process. Review cycle times have been reduced by up to 40% through the unified collaboration platform and automated checks. Checking left to manual review has decreased by 20-30%, allowing teams to focus their attention on high-value analytical tasks. Furthermore, error reduction through early discovery translates to additional 10-20% reduction in time to deliverable completion.

Efficiency gains are evident in the streamlining of various operational aspects. The elimination of redundant QC tracker updates significantly reduces administrative overhead. Communication has been streamlined, with coordination time reduced by 25% through centralized online communication platforms. The system has demonstrated scalability for larger projects, allowing organizations to handle increased workload without proportional increases in resource requirements.

Enhanced visibility provides greater insight into project status and progress. Centralized management of deliverable status enables real-time tracking and seamless collaboration across stakeholders. Executive oversight capabilities provide actionable insights and root-cause analytics to drive continuous process improvement. The capture and reporting of comprehensive audit trails supports rapid, informed decision-making through a searchable centralized record of all actions.

CONCLUSION

The future of clinical trial analysis lies not in the complete automation of processes, but in the thoughtful integration of AI capabilities with human expertise. This hybrid approach leverages the strengths of both components: the speed and consistency of Gen AI learning systems combined with the insight and adaptability of human experts. As the field continues to evolve, attention to and calibration of this balance will be crucial for ensuring both efficiency and accuracy in clinical trial analysis.

The success of this integrated approach depends on continuous refinement of both the technological tools and the human workflows that support them. Future developments should focus on enhancing the collaboration interface, growing the library of checks, and further strengthening the model that underlies the metadata curation to add greater flexibility and understanding of document and table types.

CONTACT INFORMATION

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

Author Name: Ilan Carmeli and Steve Ross
Company: Beaconcure
Address: 867 Boylston Street, 5th Floor #1491, Boston, MA
Email: Ilan@beaconcure.com; steve@beaconcure.com
Website: www.beaconcure.com
LinkedIn: http://www.linkedin.com/company/beaconcure