International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Empowering Medical Monitors: AI-Enabled Semantic Parsing for Enhanced Clinical Data Interpretation

Aditya Gadiko

Director, Product Management, Saama Technologies

Abstract: ***In the realm of clinical trials, the reliance on intricate coding for data analysis has long impeded the direct engagement of clinical professionals, such as Medical Monitors, with critical data insights. This study presents an innovative AI- based framework that utilizes semantic parsing to revolutionize data interpretation within clinical trials, effectively bridging the gap between technical execution and clinical expertise. Our approach, centered on a human-in-the-loop paradigm, enables non-technical subject matter experts to intuitively query data through natural language, which is then visualized through user- friendly UI components. This process not only encapsulates the intended inquiry but also allows for dynamic refinement by users to ensure accuracy and relevance. The introduction of this method significantly streamlines data analysis, reducing reliance on specialized programming skills and heralding a new era of data accessibility for clinical stakeholders. Through detailed case studies, we demonstrate the transformative impact of our approach, offering a glimpse into the future of clinical trials data analysis where efficiency, precision, and user engagement converge.***

Keywords: Clinical Trials, Data Analysis, Analytics, Seman- tic Parsing, Artificial intelligence, Custom Listings, No Code, Natural Language Processing, Machine Learning, AI, NLP, ML, Human in the loop, HITL

1. Introduction

a) Challenges of Data Analysis in Clinical Trials

In the realm of clinical trials, data analysis poses unique challenges that have hindered the direct involvement of clinical professionals in interpreting critical insights. Healthcare data is often standardized, fragmented, or generated in legacy IT sys- tems with incompatible formats, making it difficult to access and analyze efficiently [1]. Furthermore, the current methods in clinical trial data analysis are cumbersome, requiring in- depth knowledge of programming languages as well as an understanding of clinical trial data [2]. This creates a depen- dency on specialized resources and necessitates coordination across multiple functions, such as Bio-Statistical Program- ming, Clinical Analytics, and Data Management. As a result, these processes are resource-intensive and time-consuming, often requiring project management to handle logistics.

Accessing and understanding data poses a challenge for clinicians and non-technical staff, often requiring assistance from programming experts [3]. This dependence on coding re- sources can result in inefficiencies, potential decision-making delays, and ultimately impact patient safety. Specifically, med- ical monitors heavily depend on individuals with programming skills to extract valuable insights, causing setbacks in risk management processes and affecting patient safety.

b) The Role of AI in Transforming Clinical Trial Data Inter- pretation

AI has the potential to revolutionize clinical trial data inter-pretation by addressing the challenges mentioned above. By leveraging techniques such as semantic parsing, natural lan- guage processing, and machine learning, an AI-based frame- work can enable non-programming clinical professionals, such as Medical Monitors, to directly engage with and interpret clinical trial data [4]. This transformative approach allows for intuitive querying of data through natural language, elim- inating the need for complex coding and technical expertise. This AI-based framework utilizes semantic parsing to bridge the gap between technical execution and clinical expertise, empowering medical monitors and non-programming subject matter experts to access and interpret critical data insights in a user-friendly manner [5].

Semantic Parsing

Semantic parsing, a pivotal task in natural language processing (NLP), involves the translation of natural lan-guage utterances into logical forms—a structured, machine- comprehensible representation of meaning. This computational process is integral to applications ranging from machine trans- lation [7] and question-answering systems [6] [8] to ontology induction [9] and automated reasoning [10]. Additionally, semantic parsing plays a crucial role in the burgeoning field of code generation, enabling the conversion of natural language specifications into executable code [11] [12]. At its core, semantic parsing aims to disambiguate and precisely capture the intent and semantics of natural language, facilitating a transformative bridge between human communication and actionable, formal meaning representations suitable for ma-chine interpretation. This transduction into formal meaning representations is not restricted to a single paradigm but can be adapted to various formalisms, reflecting the versatility and depth of semantic parsing as a discipline within artificial intelligence.

Human-In-The-Loop

Human-in-the-loop (HITL) is an emergent paradigm where human feedback is incorporated into the AI model's iterative learning process. This collaboration facilitates the enhance- ment of the model's predictive accuracy and the

calibration of training outcomes. In practical machine learning applications, HITL serves as a method for humans to aid machines in making more informed decisions during the model-building phase, thereby improving learning efficiency over conventional random sampling methods [13].

Advantages of HITL include the augmentation of machine learning with human intelligence, promoting more accurate and trustworthy AI systems. It democratizes the model training process, enabling individuals without deep programming or technical expertise to engage directly with AI algorithms. Nevertheless, potential drawbacks such as the introduction of human bias and the increased requirement for human oversight can lead to escalated costs and prolonged development cy- cles [14]. Despite these challenges, HITL's approach is instru- mental in developing AI systems that are not just intelligent but also attuned to human values and ethical considerations, thus ensuring AI solutions that are effective and responsible [15]

2. Case Study: Demonstrating The Impact of AI on Clinical Trial Data Analysis

In this case study, let's pick a hurdle for a subject matter expert like a Medical Monitor, that is creating custom listings and see how can AI help democratize programming intensive task like developing custom listing across multiple domains for a clinical expert like Medical Monitors. Subject matter experts can articulate questions in natural language, which are then translated into blocks. These blocks capture the essence of the query, and if machine comprehension falters, experts can iter- atively refine the machine's understanding. The final outcome: tailored custom listings generated with precision. This method accelerates custom listing development process by many folds and fosters a more intuitive interaction with data. This case study spotlights the efficacy and the transformative potential of this methodology in real-world scenarios. Dive in to witness the future of clinical trials data analysis.

In the context of this case study, we address a significant challenge faced by subject matter experts such as Medical Monitors—the generation of custom data listings. These professionals are heavily dependent on programmers for the development of custom listings, which is traditionally a programming-intensive task spanning various domains within clinical trials. AI stands as a democratizing force, transforming complex programming tasks into a more accessible, language- driven paradigm.

Medical Monitors can formulate queries using natural language, which this AI-based framework subsequently deconstructs into discrete computational blocks. These blocks serve as interpretable and logical representations of the queries, ensuring that the core intent is preserved. Should there be a discrepancy in machine interpretation, the iterative nature of this method permits experts to refine their input, thereby guiding the machine toward an improved understanding. The resultant listings are not only custom-tailored with high precision but are also produced with considerable speed compared to traditional methods.

The methodology unfolds in three structured steps, utilizing natural language processing to facilitate the generation of validated data listings for clinical experts such as Medical Monitors.

In Step 1, a natural language query is inputted, as presented in Figure 1 "Show me subjects with hypertension as AE, CM with Diphenhydramine, age greater than 60." This step comprises four substeps:

1) Autocompletion: As the user composes the query, the AI assists by recommending words, expediting the query formulation, and minimizing typographical errors.
2) Entity Recognition: The AI identifies key entities within the query, singling out terms of specific relevance to the clinical context.
3) Data Type Classification: Recognized entities are then classified into appropriate categories—categorical, numerical, or date-based—according to the data type they represent.
4) Data Model Mapping: The classified entities are systematically mapped to a predefined data model, ensuring coherence with the established database schema.

Step 2 involves semantic parsing, a pivotal transformation where the machine interprets the natural language input into a logical format understandable by the AI system. This stage serves a dual purpose: validating the AI's interpretation of the query and allowing users to adjust the logic through a user- friendly interface if discrepancies arise.

The process culminates in Step 3 with the generation of the final output. Upon user approval of the logical constructs presented in Step 2, a confirmation action yields the custom data listing. The system also accommodates feedback on this output, enabling further refinement for subsequent queries.

This sequential methodology not only streamlines the cre- ation of custom listings but also embodies a more interactive and intuitive form of programming, marking a substantial advancement from traditional, code-intensive practices. The case study underscores the robustness and transformational capability of this AI-driven approach in real-world clinical trial data analysis scenarios.

The protocol for refining AI comprehension of clinical data encompasses numerous iterative user interactions, extending beyond the initial natural language query formulation.

As delineated in Figure 2, user engagement is not con- strained to the logic encapsulated by the natural language query crafted in Step 1. Users can enhance the query's com- plexity and specificity by engaging with the live database—a feature signified by the plus icon, emphasized in yellow within Figure 2. This interaction grants the user direct access to the database schema, empowering them to seamlessly integrate additional logic into their query by selecting relevant tables, columns, and subcategories that exist within the live database. This dynamic feature

obviates the need for guesswork regarding the database's contents, enabling users to construct more robust and data-driven logical blocks. By leveraging this functionality, users can augment the predefined logic with actual data structures, fostering a more robust and accurate representation of their intended analysis. Such capabilities are instrumental in formulating queries that are not only syntactically correct but semantically rich and closely aligned with the underlying data model, thereby enhancing the precision of the resultant data listings.

Within the user interface for refining AI interpretation of queries, an essential feature is the capacity to modify

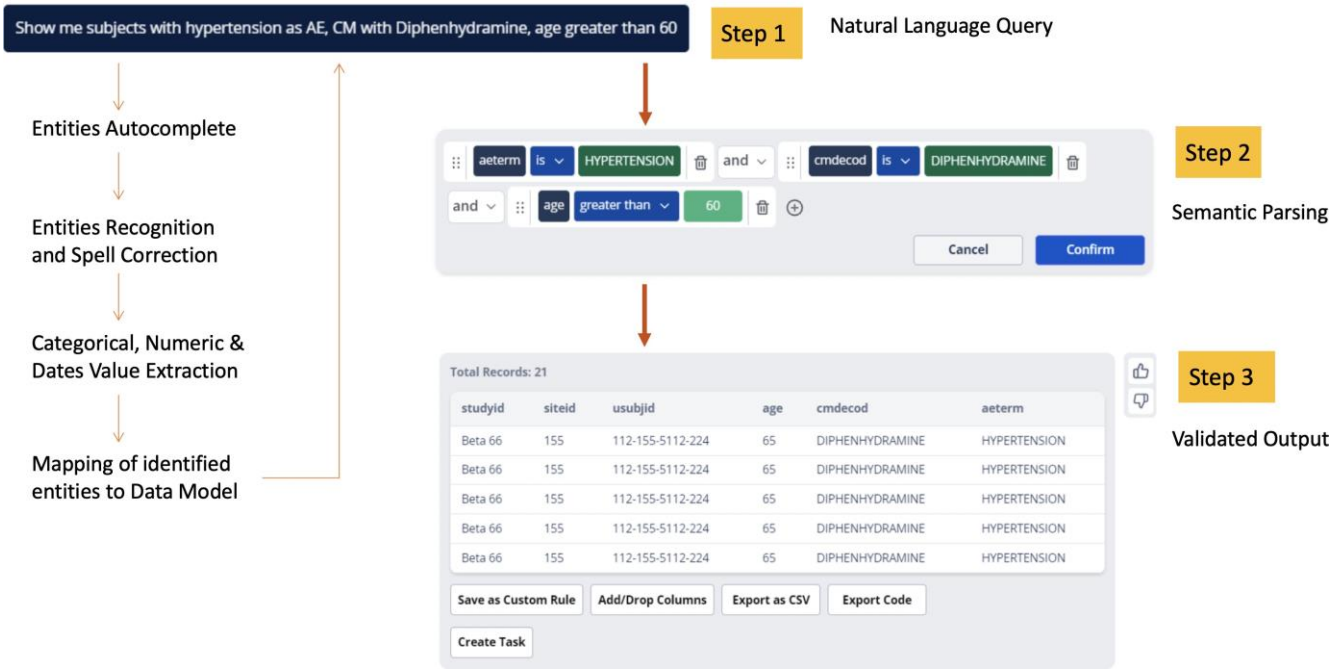

Figure 1

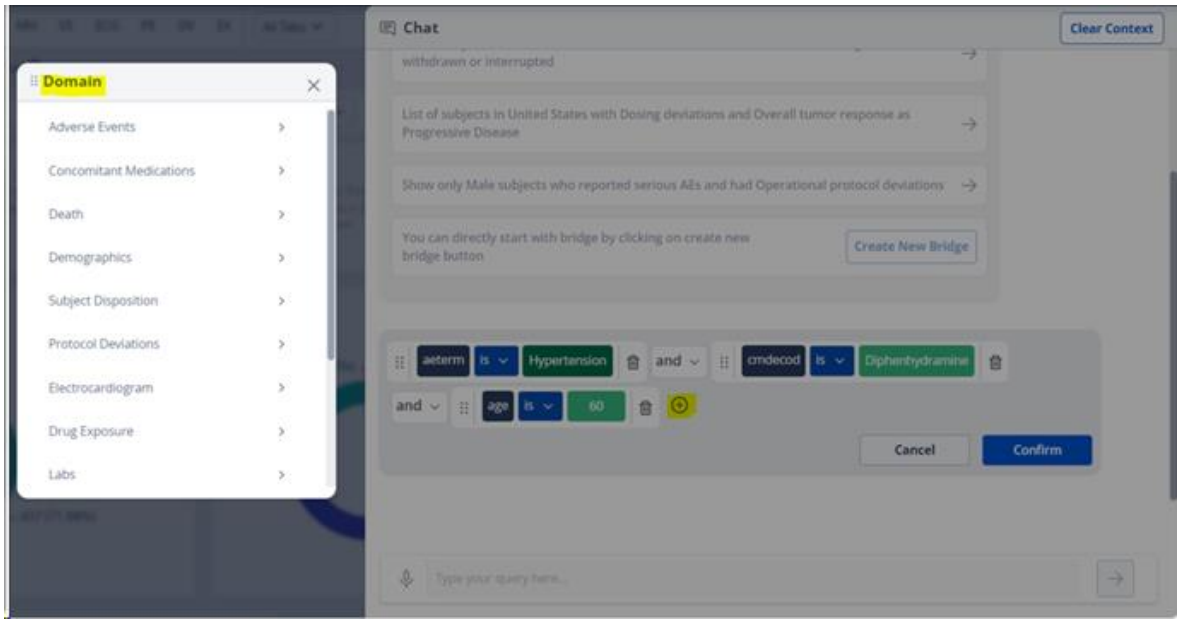

Figure 2

Figure 3

logical connectors, enhancing the complexity of the conditions applied. As illustrated in Figure 3, users possess the flexibility to alternate between logical operators such as 'And' and 'Or.' This functionality is particularly pivotal when formulating compound conditional statements that necessitate grouping. The ability to switch operators facilitates the construction of more nuanced and layered logical expressions, especially when handling multiple conditions that require distinct groupings for accurate data retrieval.

The availability of this logical editing feature underscores the system's adaptability, permitting users to fine-tune the query parameters in real-time and ensure that the resulting data set is a true reflection of the complex clinical scenarios being investigated. Such granular control over the logic enables clinicians and researchers to craft precise and contextually rel- evant queries, thus yielding data listings that are meticulously aligned with their investigative criteria.

The system's design facilitates an intuitive interface that allows users to manipulate and structure complex query conditions through a drag-and-drop functionality for grouping as shown in Figure 4. This becomes particularly potent when paired with the logical 'And'/'Or' operators, thereby empowering non-programmers, such as Medical Monitors, to intuitively construct grouped conditions across multiple data domains.

In the specialized context of safety monitoring, for instance, Medical Monitors are often tasked with the simultaneous review of patient data across disparate tables—namely Adverse Events, Exposure, Labs, and Concomitant Medications. The drag-and-drop feature, combined with logical connectors, equips them to seamlessly aggregate and cross-reference data points pertaining to individual patients from these varied tables.

Illustrated in Figure 5, users can engage with a discrete subcategory—take 'Hypertension' as an example—by sim- ply hovering over it and initiating the edit function. This action reveals a spectrum of available alternatives within the 'AETERM' column sourced directly from the live database.

Such a feature endows Medical Monitors with the flexibility to not only select from existing sub-categories but also to tailor these categories subsequent to formulating a natural language query. The ability to dynamically interact with and modify sub-categories post-query creation enhances the precision of search results and allows for the curation of data reflective of the intricacies inherent in clinical monitoring scenarios. This exemplifies the system's capability to adapt query parameters in real-time, thereby supporting Medical Monitors in their objective to conduct granular and focused analyses.

The system's utility is further enhanced by the capacity to manipulate logic for both categorical and numeric data types, as exemplified in Figure 6. Users can employ a suite of numerically relevant operators including 'greater than', 'less than', 'greater than or equal to', 'less than or equal to', 'is', 'between', and 'is not'.

This array of functions grants users comprehensive control over the construction of numeric logic, crucial for delineating precise conditions within their analytical queries. By integrating such capabilities, the interface empowers users to refine and sculpt the logic to match specific numerical criteria, leading to outputs that are meticulously tailored and highly relevant to the investigation at hand.

The inclusion of this functionality is particularly significant in clinical data analysis, where numeric values often carry critical information pertaining to patient outcomes, Lab values, Vital Signs, treatment efficacy, and other vital statistics. Thus, the ability to accurately define and apply numeric conditions is imperative for the extraction of meaningful insights, and this system adeptly furnishes Medical Monitors with the tools necessary for such sophisticated data manipulation.

Users will be able to switch between 'is' and 'is not' for all data types. As shown in Figure 7, users can create an exception listing by just choosing 'is not' via which they will be able to generate a listing of all patients who have not used 'Diphenhydramine'

The analytical capabilities of the system are significantly expanded by the inclusion of temporal logic construction. As demonstrated in Figure 8, the interface empowers users to incorporate date-based criteria into their queries, facilitating the examination of temporal data with precision.

Available temporal operators include 'between', 'greater than', 'less than', 'greater than or equal to', 'less than or equal to', 'is', and 'is not'. Such functionality is indispensable for users who require the application of time-bound conditions in their data review, such as defining a cut-off date for data extraction. This temporal dimension is crucial when analyzing longitudinal data or when outcomes must be assessed within specific time frames.

The adept handling of date parameters within the logic-building process enables users to pinpoint data that is tempo- rally relevant, providing the means to conduct rigorous, time- sensitive reviews. While the mentioned application pertains

Figure 4

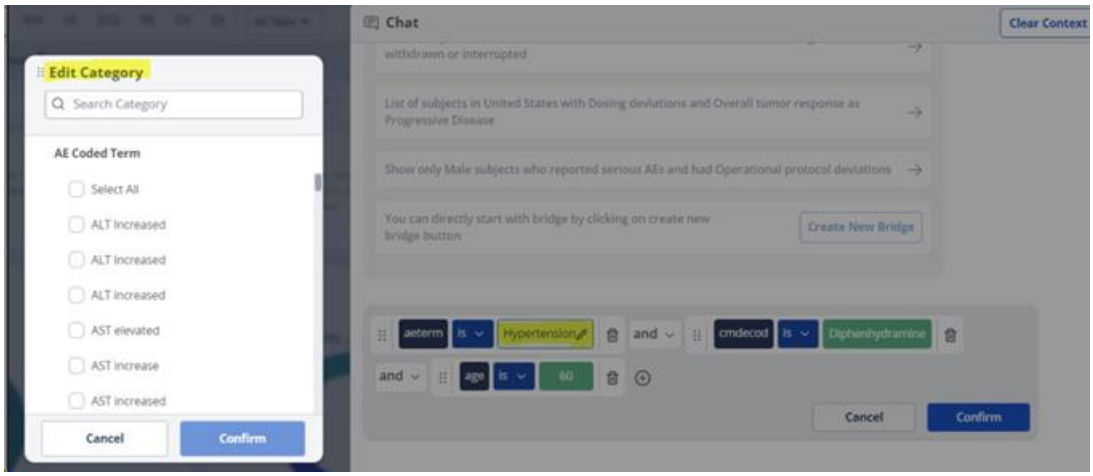

Figure 5

Figure 6

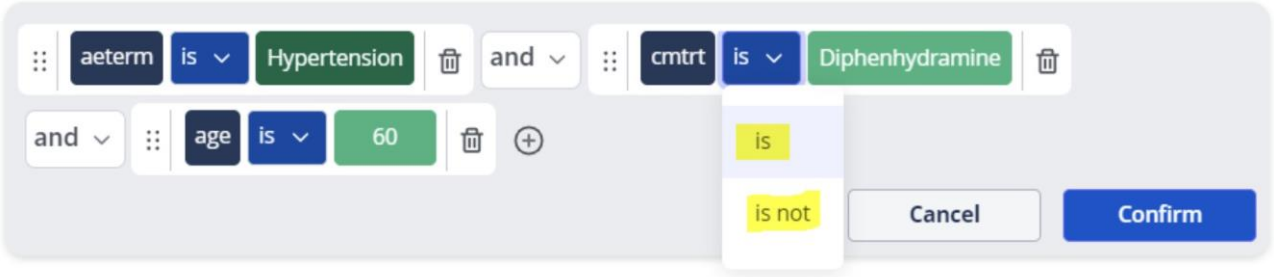

Figure 7

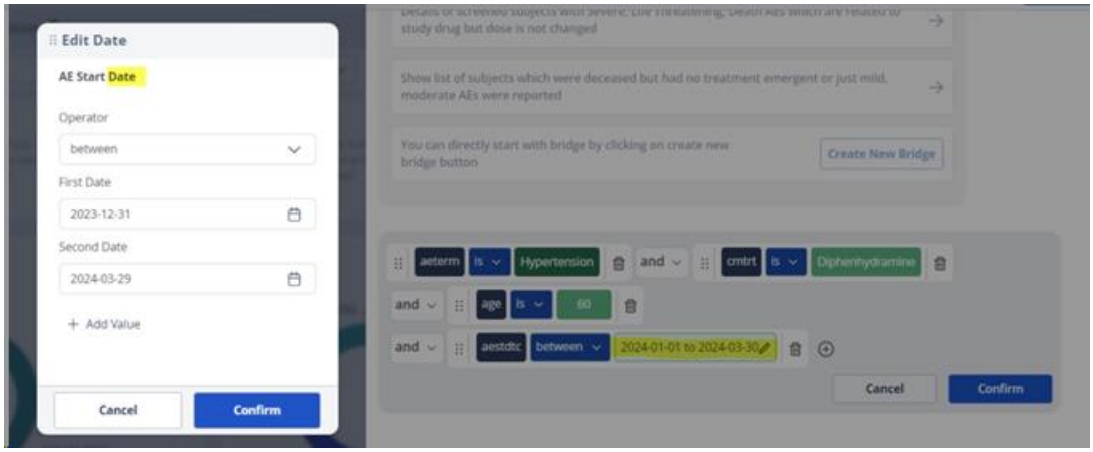

Figure 8

to cut-off date analysis, the system's versatility extends to a myriad of other scenarios where time-based data is a pivotal factor in clinical assessment and decision-making processes.

3. Conclusion

The empirical evidence presented in the case study vividly illustrates the radical impact of AI on the analytics of clinical trial data. The proficiency of AI to swiftly interpret and execute complex queries, culminating in the production of validated data listings, represents a paradigmatic shift away from conventional methodologies that typically demand an extensive temporal investment.

Such a progression in data analytical approaches is indicative of an imminent era characterized by augmented adaptability and reactivity within clinical trials [16]. The expediency and precision inherent in AI capabilities have the potential to expedite critical interventions, bolster patient safety, and streamline the management of clinical trials, particularly those of a larger scale or dispersed across multiple sites, where the intricacy and sheer quantity of data present formidable challenges [17].

Nevertheless, the AI model in its current form is not without its shortcomings. Presently, it lacks the capability to perform quantitative count operations or conduct intricate statistical analyses—functions that are pivotal for a holistic interpretation of data. These enhancements are anticipated in forthcoming iterations of the system. Concurrently, the integration of AI into established clinical trial frameworks raises several concerns. Among these, the preservation of data confidentiality and security, adherence to stringent regulatory standards, and seamless technological integration stand as critical issues that necessitate careful navigation as this technological evolution progresses [18] [19].

In conclusion, this paper has introduced a pioneering artificial intelligence (AI) methodology that represents a quantum shift in clinical trial data analysis, discarding the cumbersome code-dependent techniques of yesteryears. Central to this innovation is the empowerment of Clinical Experts to engage with the AI through natural language inputs. The resulting 'Blocks' encapsulate the Clinical Experts' intents, thus streamlining and enhancing the fidelity of data interpretation. Through iterative Clinical Expert-engaged refinements, the AI's interpretations become intricately attuned to the expert's discernment, bolstering the trustworthiness of the generated outputs. Our case studies underscore the profound efficiency gains enabled by this system, which condenses listing generation to a matter of hours—a drastic reduction from the traditional spans of weeks—thereby expediting critical decision-making in clinical trial operations.

This AI-driven method heralds a new era in clinical trial data analysis by fusing clinical acumen with sophisticated technical processing, affording a more intuitive and dynamic interaction with trial data. Such progress not only promises speedier analytical procedures but also ensures the accuracy and reliability of the findings, factors vital to ensuring patient safety and evaluating treatment efficacies. As the horizon of AI in clinical trials expands, forthcoming research and development endeavors are poised to augment the current model with advanced functionalities such as count operations and comprehensive statistical analysis. Concomitantly, as AI's integration into clinical trials deepens, resolving challenges linked to data privacy, regulatory adherence, and harmonization with existing infrastructures will be paramount. The objective is to cultivate an environment where clinical trial data analysis is not only more adaptable and responsive but also more efficient, thereby catalyzing breakthroughs in medical research and enhancing patient outcomes

References

[1] T. Wilkinson, S. Sinha, N. Peek, and N. Geifman, "Clinical trial data reuse - overcoming complexities in trial design and data sharing," *Trials*, vol. 20, no. 1, 2019.
[2] B. Krishnankutty, S. Bellary, B. N. Kumar, and L. S. Moodahadu, "Data management in clinical research: An overview," *Indian Journal of Pharmacology*, vol. 44, no. 2, pp. 168–168, 2012.
[3] S. Denaxas, F. W. Asselbergs, and J. H. Moore, "The tip of the iceberg: challenges of accessing hospital electronic health record data for biological data mining," *BioData Mining*, vol. 9, no. 1, 2016.
[4] K. Kreimeyer. A. Wen, "Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation," *npj Digital Medicine*, vol. 2, 2019.

[5] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on Free- base from question-answer pairs," *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1533–1544, 2013.
[6] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," *Proceedings of the 2003 Conference of the North American Chapter*, vol. 1, pp. 48–54, 2003.
[7] M. Richardson, C. J. C. Burges, and E. Renshaw, "MCTest: A Chal- lenge Dataset for the Open-Domain Machine Comprehension of Text," *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 193–203, 2013.
[8] H. Poon and P. Domingos, "Joint unsupervised learning of semantic roles and prototypical event schemas," *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 459–468, 2010.
[9] E. Zelle and R. J. Mooney, "Learning to parse database queries using inductive logic programming," *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, vol. 1, pp. 1050–1055, 1996.
[10] R. Le and T. A. L. Nguyen, "From natural language to code: Leveraging transformer models for code generation," *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 1–6, 2020.
[11] Y. Ling, S. R. Bowman, S. Shrivastava, G. Ilharco, J. Gauthier, and V. Chaudhary, "CoNaLa: The Code/Natural Language Challenge," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, pp. 369–375, 2019.
[12] V. S. Bisen, "What is Human in the Loop Machine Learning: Why & How Used in AI?" *Medium*, 2020.
[13] C. Chandler, P. W. Foltz, and B. Elvevåg, "Improving the Applicability of AI for Psychiatric Applications through Human-in-the-loop Method- ologies," *Schizophrenia Bulletin*, vol. 48, no. 5, pp. 949–956, 2022.
[14] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes- Bascarán, and A. Fernández-Leal, "Human-in-the-loop machine learn- ing: a state of the art," *Artificial Intelligence Review*, vol. 56, pp. 3005–3054, 2023
[15] S. Askin, D. Burkhalter, G. Calado, E. Dakrouni, and S, *Artificial Intelligence Applied to clinical trials: opportunities and challenges. Health Technol (Berl)*, vol. 13, no. 2, pp. 9 974 218–9 974 218, 2023.
[16] H. Chopra, Annu, D. K. Shin, K. Munjal, Priyanka, K. Dhama, and T. B. Emran, "Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs," *Int J Surg*, vol. 109, no. 12, pp. 10 720 846– 10 720 846, 2023.
[17] M. V. Wickramasinghe and D. S. Wishart, "Artificial Intelligence and Machine Learning in Clinical Trial Data Science: Opportunities and Challenges," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 652–663, 2020.
[18] E. J. Topol, "The Future of Clinical Trials: Adaptive, Personalized, and Accelerated Designs Enabled by Big Data and AI," *Annual Review of Medicine*, vol. 70, pp. 1–13, 2019.