Paper ML25

GPT-Powered Data Insights: Advancing Clinical Data Analytics with Large Language Models (LLMs)

Melanie Hullings, Formation Bio, New York, NY, US

ABSTRACT

This presentation explores how Large Language Models (LLMs) can revolutionize data analytics. We demonstrate LLMs' capabilities in integrating and analyzing information from diverse formats. Our focus is on leveraging GPT technology to enhance clinical data exploration and interpretation, facilitating rapid knowledge extraction, visualization, analysis, and data-driven decision-making. We showcase approaches for automated data processing, interactive analysis, and advanced visualization techniques. Crucially, we address the validation of AI-generated insights, presenting methodologies for ensuring result accuracy and reliability. This includes comparative analysis with traditional statistical methods, domain expert verification, and the implementation of explainable AI techniques. Our innovative approach streamlines analytical processes, improves data interpretation, and uncovers quick insights in clinical research, contributing to more efficient drug development and improved patient outcomes.

INTRODUCTION

WHY LLMS IN CLINICAL DATA ANALYTICS?

The exponential growth of clinical trial data presents significant challenges in data processing, standardization, and validation. Traditional, rule-based methods struggle to manage the complexity and volume of modern clinical datasets, leading to inefficiencies in data analysis and clinical trial execution. LLMs, such as GPT-4 and even newer models emerging, offer a novel solution by enabling intelligent automation to augment traditional programming methods in clinical data analytics.

Much like geneticists decode DNA to understand biological functions, LLMs decode clinical data to extract meaningful insights while addressing common challenges such as:

- **Data Growth & Complexity:** The increasing volume of structured and unstructured clinical data makes manual analysis unscalable.
- **Data Quality:** Ensuring clean, standardized datasets for analysis is labor-intensive and error-prone.
- **Regulatory Requirements:** The FDA and EMA mandate structured, traceable, and validated data reporting.[1]

LLMs provide a scalable approach to these challenges by automating data processing and facilitating advanced analytics, ultimately improving efficiency and reducing time-to-insight.

Additionally, as clinical research continues to generate massive amounts of data, organizations must leverage AI-driven tools to manage diverse data types from various sources, including electronic health records (EHRs), adverse event reports, and patient-reported outcomes. LLMs have the potential to bridge the gap between unstructured narratives and structured datasets, ensuring a more comprehensive view of trial outcomes and patient safety profiles.

TRANSFORMING CLINICAL DATA ANALYTICS WITH LLMS

There are three specific capabilities of LLMs that can be applied across clinical data review workflows to augment manual reviews of text, data, and data trends.

NATURAL LANGUAGE UNDERSTANDING (NLU)

LLMs are designed to interpret and generate human-like text, making them particularly effective in extracting insights from unstructured clinical narratives. For example, AI-powered models can scan investigator comments, protocol amendments, and medical notes to highlight key risk factors, treatment deviations, and emerging safety concerns. This capability not only improves real-time decision-making but also reduces the burden on medical reviewers by summarizing large volumes of text efficiently.

PATTERN RECOGNITION IN STRUCTURED DATA

Beyond unstructured narratives, LLMs can analyze structured datasets, identifying anomalies and inconsistencies that may go unnoticed through traditional rule-based checks. By applying probabilistic reasoning, these models can detect missing data patterns, assess correlations between treatment responses, and flag deviations from expected

patient progressions. Moreover, integrating LLMs with statistical models enhances the precision of outlier detection, reducing false positives in data monitoring.

AUTOMATED REPORT GENERATION

One of the key applications of LLMs in clinical analytics is automated report generation. AI models can create draft reports for clinical study results, adverse event trends, and safety summaries, reducing the time required for regulatory submissions. These models can also generate structured narratives for Data Monitoring Committees (DMCs), ensuring consistency across trial sites and facilitating data-driven decision-making. By automating repetitive aspects of clinical reporting, LLMs allow data scientists and medical reviewers to focus on interpreting findings and validating conclusions.

A recent study demonstrated that GPT-4 outperformed traditional rule-based NLP methods in summarizing unstructured EHR data and extracting clinical phenotypes.[2] The performance of GPT-4 and 6 other LLM models were compared to gold-standard manual annotation from 2 subject matter experts. GPT-4 demonstrated data-driven effectiveness even with limited context in the input, improved contextual understanding of the text, and robust clinical phenotype extraction.

It is important to note that in this context, LLMs are not replacing any of the reviews performed by subject matter experts, but are augmenting their workflows to make review more efficient but do require feedback and confirmation from clinical reviewers.

GPT-POWERED WORKFLOWS IN CLINICAL DATA ANALYTICS

If clinical data were DNA, then we—data analysts—would be geneticists trying to decode, repair, and interpret it. But just like DNA, clinical data is noisy, messy, and full of mutations. AI serves as a data geneticist that can help us clean, structure, and analyze this data faster. LLMs act as the geneticist of clinical data analytics—edit, interpret, and assist in validating insights for better decision-making. We'll walk through how to use AI, specifically ChatGPT, in these examples to translate data DNA into high quality proteins for the cell (aka actionable data insights).

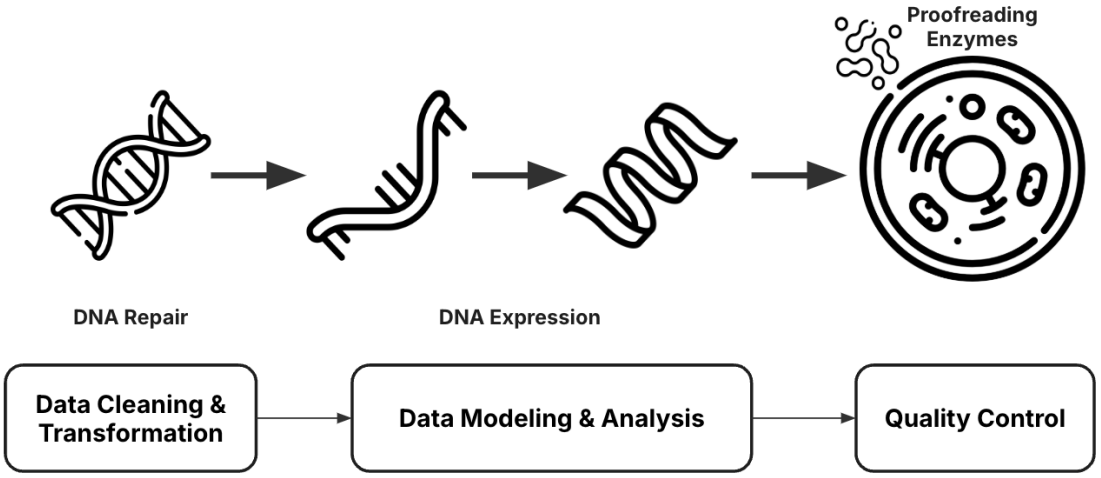

Figure 1: Parallels in DNA and data processing and interpretation workflows.

LLMs support the end-to-end data analysis workflow from data cleaning, transformation, modeling, and analysis, in addition to aiding in QC of generated outputs. Though initially best at language interpretation, recent advancements have significantly enhanced the mathematical and quantitative reasoning capabilities of LLMs. OpenAI's o3-mini model, released in January 2025, was designed as a cost-effective and efficient model that excels in science, technology, engineering, and mathematics (STEM) tasks, with notable strengths in mathematics and coding.[3]

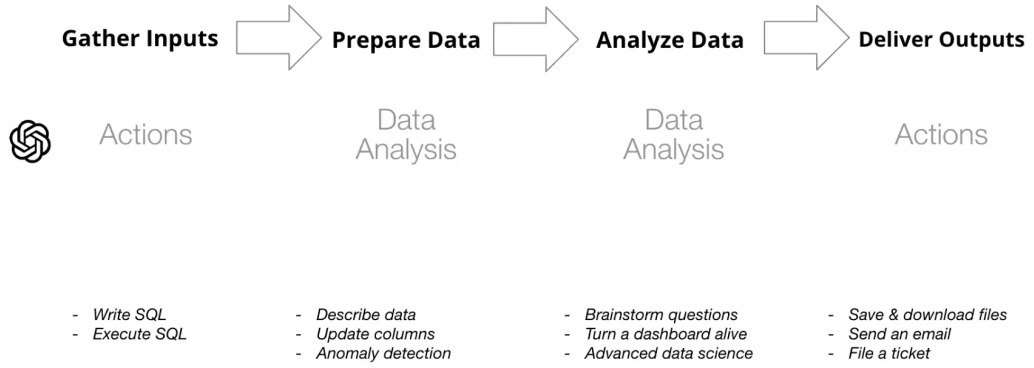

Figure 3: GPT-supported end-to-end data analysis workflow.[4]

Many AI-powered data analytics platforms are now easily accessible to users and are very simple to use. For example, OpenAI's Data Analyzer Custom GPT is a specialized tool within the ChatGPT platform designed to facilitate advanced data analysis through natural language interactions. Leveraging the capabilities of GPT-4, this tool allows users support in querying databases using SQL code if needed. Users can then upload datasets directly into the chat interface and pose questions or requests in plain English. The model then interprets these prompts, generates the necessary Python code, executes it in a secure environment, and returns the results, which can include data visualizations and insights.[5]

Our team performed a comprehensive analysis of the Osteoarthritis Initiative (OAI) study's publicly available database.[6] The OAI study is a comprehensive, multi-center, longitudinal observational study aimed at enhancing the understanding of knee osteoarthritis (OA) development, progression, and prevention. Sponsored by the National Institutes of Health (NIH), the OAI has established a rich public-domain data resource facilitating the scientific evaluation of biomarkers as potential surrogate endpoints for OA.

Below are examples of how we used AI to support this analysis throughout the process with a variety of methods and tools. But first, we want to

1. Data Cleaning & Transformation (DNA Repair)

LLMs enhance data integrity by automating the detection of inconsistencies and formatting errors in clinical datasets. For example, natural language processing (NLP) techniques can infer missing values based on contextual data, reducing the need for manual imputation. Additionally, LLMs can enhance understanding and exploration of a dataset to identify key fields for analysis and perform data transformation. These improvements lead to cleaner datasets that are more suitable for downstream statistical analysis.

Example Application: Basic prompting allows for AI to make suggestions on data cleaning checks and summarize data quality, automatically flagging missing variables and suggesting imputation strategies based on available historical trial data. As this data resource also provides extensive documentation including data dictionaries, we also developed a specific custom GPT model that aggregated information across all file types and allowed us to provide very specific prompts to perform data transformations to produce analysis-ready tables as output. AI significantly reduced manual review time while maintaining data integrity.

2. Data Modeling & Analysis (Gene Expression)

In data summarization and visualization, LLMs assist in interpreting clinical questions and provide tables with population metrics, as well as generate charts and graphs. In addition, AI can help to generate complex R code to create Shiny applications that allow researchers to interact with complex datasets through intuitive visualizations, enhancing their ability to identify clinically significant trends. Furthermore, LLMs can automate the selection of statistical models by analyzing study design parameters, optimizing the accuracy of trial endpoints while reducing human bias.

Example Application: LLMs supported the statistical analysis plan of the OAI database, as well as development of an R Shiny app to interactively explore patient data. The LLM assisted in writing R code to make the app and analyses fully dynamic so that researchers could filter the data in many ways to create sub-populations by endpoint and receive instant data visualizations, making iterative analysis more accessible to non-programmers.

3. QC & VALIDATION OF AI-GENERATED OUTPUTS (PROOFREADING ENZYMES)

Regardless of the classification of data, critical thinking is required to evaluate risk based on the data, information, and decisions being made for each use case, including the ability to quality control (QC) outputs because AI can be inconsistent, factually incorrect, and hallucinate. Always remember to review and QC output to ensure it is accurate before drawing conclusions or making decisions.

Example Applications:

- **Logical checks:** Risk-based, basic spot checking focusing on critical data and analyses that drive underlying assumptions and decisions.
- **Ask for sources:** Ask ChatGPT to explain the steps taken to generate an analysis and provide equations and/or references.
- **Output verification (i.e., human-in-the-loop):** Expert review of AI-generated insights to ensure accuracy and reliability.
- **Awareness of bias in the data:** Consider how the dataset could be limited or biased, and proactively evaluate the potential impact (ask GPT for bias analysis recommendations).

KEY TAKEAWAYS

➢ **LLMs enhance data processing efficiency.** By automating data cleaning, transformation, and modeling, LLMs significantly reduce the manual effort required in clinical research. This efficiency allows data scientists to focus on higher-value tasks such as interpretation and strategic analysis, ultimately accelerating drug development timelines.

➢ **AI-powered analytics improve data quality and accessibility.** LLMs facilitate data standardization and error detection, leading to higher data integrity. Furthermore, AI models can extract meaningful insights from both structured and unstructured data, making clinical datasets more interpretable for diverse stakeholders, including regulatory bodies and research teams.

➢ **Shiny apps with LLM integration enable real-time interaction with clinical datasets.** These applications allow users to perform dynamic queries, generate interactive visualizations, and automate exploratory data analysis. This capability enhances decision-making by providing on-demand insights that adapt to evolving research needs.

➢ **Validation and QC of AI-generated outputs remain critical.** While LLMs enhance clinical data workflows, their outputs must undergo rigorous validation to ensure accuracy and regulatory compliance. Implementing human-in-the-loop verification, bias detection, and consistency checks are essential to maintaining trust in AI-driven analytics.

FUTURE DIRECTIONS

➢ **Advancements in Multi-Modal AI.** Future models like GPT-5 and Med-PaLM 2 will integrate text, images, and structured datasets, enabling deeper analysis and broader applications in clinical trials. These models will provide richer, context-aware insights, particularly in multi-source data environments where combining clinical notes, imaging results, and tabular data is necessary.

➢ **Regulatory Landscape Evolution.** As AI adoption grows in clinical research, agencies such as the FDA, EMA, and MHRA will refine guidelines to address AI transparency, validation, and ethical considerations. New standards will likely emerge to ensure compliance while enabling the innovative use of AI in pharmaceutical development.

➢ **Ethical and Interpretability Challenges.** AI explainability remains a key concern, especially in regulated environments where clinical decisions rely on data-driven insights. Efforts to improve model transparency, minimize algorithmic bias, and create interpretable AI outputs will be essential to fostering trust and broader adoption in clinical research.

These developments will shape the future of clinical data analytics, driving innovation while ensuring that AI-generated insights remain reliable, interpretable, and aligned with regulatory expectations.

REFERENCES

1. FDA Good Machine Learning Practices for Medical Device Development. U.S. Food and Drug Administration (FDA). October 2021.
2. Bhattarai et al. (2024). LLMs for Predictive Analytics & Pharmacovigilance in Clinical Trials.
3. OpenAI. (2025, January 31). OpenAI o3-mini. OpenAI. https://openai.com/index/openai-o3-mini/
4. OpenAI (2024). OpenAI Presents AI-Empowered Data Analysis.

5. OpenAI. (2024, July 15). Data analysis with ChatGPT. OpenAI Help Center. https://help.openai.com/en/articles/8437071-data-analysis-with-chatgpt
6. Osteoarthritis Initiative. (2025). Osteoarthritis Initiative (OAI) Database. National Institute of Mental Health Data Archive. https://nda.nih.gov/oai

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Melanie Hullings
Company: Formation Bio
Email: melanie@formation.bio