Computer Fraud and Security
ISSN (online): 1873-7056

Zero-Shot Clinical Trial Design Using Large Language Models: A Data-Driven Approach to Protocol Generation and Cohort Identification

Amarnath Reddy Kallam

Senior Manager & Solution Architect

Abstract— This project examines the innovative use of Large Language Models (LLMs) in designing zero-shot clinical trials in oncology. The method permits autocomplete of the entire trial protocol based on limited information-name of the drug, mechanism, indication, and phase- without speciality training-related tasks. The context supplied by historical Sanofi-financed clinical trial shows relative background to what the model yields. It involves the generation of protocols, the extraction of inclusion/exclusion criteria, the design of pseudo-SQL queries to find the cohort of patients, and the creation of regulatory submission documents. The research showed that LLMs are capable of independently and effectively generating quality trial documentation. This zero-shot approach has the potential to save time and cost of initiating a trial and increase inclusivity by appropriately choosing cohorts solely based on characteristics of real-world data.

Keywords— *Zero-Shot Learning, Large Language Models (LLMs), Clinical Trial Protocol Generation, Biomedical Natural Language Processing, AI in Drug Development*

I. INTRODUCTION

The Traditional clinical trial design is one of the most cost-capital and time-consuming phases in coming up with new medical processes. It needs interdisciplinary cooperation, space-authorized regulation and expertise. Clinical procedures are especially challenging to design because they must strike a balance between ethical considerations, scientific priorities and patient safety. These characteristics of a continually evolving protocol development process are often associated with slow and expensive outcomes due to lack of historical analogs and changing stakeholder requirements. Such issues are further exacerbated by weaknesses in choosing representative cohorts and optimising designs in diverse populations.

The development of Natural Language Processing (NLP), particularly Large Language Models (LLMs), presents a unique opportunity in biomedical research. Currently, LLMs can comprehend, summarise, and produce high-quality scientific materials based on large medical and trial databases. The project suggests the use of an LLM in zero-shot clinical trial design,

allowing the comprehensive automatic generation of trial documentation without domain-specific retraining.

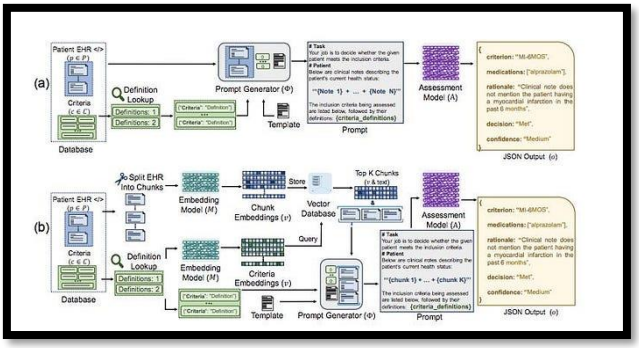


Fig 1**:** End-to-End Workflow of Zero-Shot Clinical Trial Design Using LLMs

The targeted design is to apply historical clinical trial records funded by Sanofi as a background to guide the development of a Phase II protocol of an imaginary KRAS G12C blocker meant to treat metastatic colorectal cancer. The structured input supplied to the LLM consists of the properties of drugs, therapeutic targets, and trial examples. It edits eligibility criteria extracts a complete trial protocol, creates pseudo-database queries to identify patient cohorts in databases, and a regulatory summary that can be included in an Investigational New Drug application. The general objective is to determine the extent to which a general-purpose LLM makes end-to-end trial planning zero-shot.

II. LITERATURE REVIEW

The Artificial Intelligence (AI) revolution has transformed clinical trials development, data analysis, and identification of medical insights in biomedical research. Statistical learning and rule-based systems were among the earliest applications of AI to clinical practice; more recently, widespread manipulation of unstructured medical data has become possible with deep learning. Specifically, natural language processing (NLP) has become central in the study of clinical narratives, electronic health records, and trial protocols [1]. Such domain-specific models as BioBERT and ClinicalBERT enable the processing of medical texts with a high level of refinement, to recognize clinical concepts, assess patient eligibility and identify adverse events. Closely related to Gibbon, systems such as IBM Watson Health also demonstrated how AI can support diagnosis and aid in selecting treatment but were limited in their flexibility and cross-domain generalisation.

Recent developments in Large Language Models (LLMs) have broadened the realm of AI in biomedical applications. Research indicates that finely tuned LLMs can encode and group clinical eligibility criteria, summarise protocols without forfeiting clinical significance, and normalise ideas in a zero-shot fashion through alignment of variable clinical language with standardised medical terminologies [2]. This minimises the use of task-specific training data, increasing the applicability of medical NLP.

Nonetheless, existing models would still use much of the fine-tuning and have difficulty creating full trial protocols using only a small amount of input. Accordingly, there is still a distance between a trial concept and its implementation. The methodology proposed in this project helps to fill this gap in creating a full Phase II oncology trial using a general-purpose LLM model in zero-shot setting. Based on historical trial data, the protocol, inclusion and exclusion criteria, pseudo-database queries to identify patients, and regulatory documents will be generated by the LLM [3]. This is a major change compared to the current application of AI that focuses on data extraction, not independent scientific and regulatory design.

III. DATASET AND PREPARATION

The data utilized in this paper is a targeted subset of sponsored clinical trials by Sanofi companies which are identified through a bandwidth constrained stream of publicly accessible trial metadata. It holds more than 3,000 records, each is a record of clinical trial which is marked with metadata, including NCT number, sponsor, study title and summary start year/month, study stage, enrolment, status, and medical condition. Though of varying therapeutic focus, the dataset is highly weighted towards phase III trials and contains fewer early entries.

A filtering system was used to find trials associated with Colorectal Neoplasms to ensure that the research was specific to metastatic colorectal cancer. This provided a limited number (two phase III studies, which have formed the contextual foundations in activating the Large Language Model [1]). These trials were chosen due to relevance, objective understanding, thorough treatment description, and methodology.

One major weakness caused by the late-phase dominance of the dataset was the absence of representative Phase II trials- incredibly significant to the design mission. Hypothetical extractions were consequently performed to estimate Phase III parameters (e.g., sample size and endpoints) into a Phase II setting. Further processing involved work with null values, phase formatting problems, and exclusions of trials with insufficient summary information.

IV. METHODOLOGY: ZERO-SHOT PROMPT ENGINEERING

This study uses the methodology of zero-shot prompt engineering to generate a clinical trial protocol using a general-purpose Large Language Model (LLM) in an autonomous fashion. The methodology in focus was to develop a very structured and informative prompt that would resemble the input that an expert clinical researcher would give when planning a new trial. Instead of fine-tuning or retraining, it was hoped that, based on the thorough prompt engineering practices and anchoring the prompt to real life, the model would produce correct and contextually relevant outputs.

The format of the prompt was logical and sequential [3]. It began by socializing a professional identity onto the LLM, making it a specialist in clinical trial design with expertise in oncology and regulatory systems. This role anchoring was subsequently accompanied by specification of a task: to create an inclusive clinical trial protocol about a new investigational drug. Then key parameters were presented, like the name of the hypothetical drug (NewDrugX), its mechanism of action (oral KRAS G12C inhibitor), target indication (metastatic colorectal cancer), and the trial phase to be entered (Phase II). These were presented in a systematic manner to make them easy to understand and relevant.

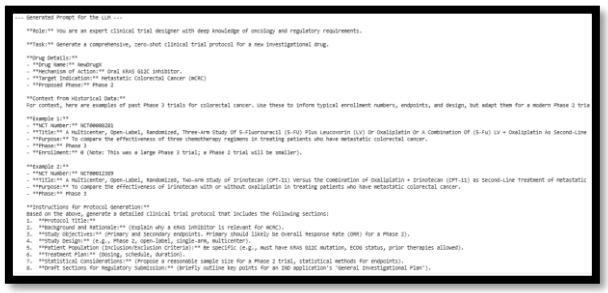


Fig 2: Prompt Structure with Contextual Grounding from Historical Trial Data

To enhance context, two historical trials related to metastatic colorectal cancer were taken from the filtered Sanofi-sponsored dataset and incorporated into the prompt. The metadata of these examples included NCT number, study title, objective of the trial, phase classification [4], and patient enrolment numbers. The prompt grounded the model through real-world trial examples so that it could infer common design principles and adapt them to the Phase II setting.

The model was requested for the generation of specific protocol sections, namely, study title, background and rationale, study objectives, trial design, eligibility criteria, treatment plan, statistical framework [6] and a summary for regulatory submission. The system was able to produce a comprehensive clinical trial design in response to a prompt, reflecting an internal coherence among the different sections of the output.

After protocol generation, the inclusion and exclusion criteria were automatically parsed from the output using regular expressions. The extracted criteria were simulated to create a pseudo-SQL query to identify the appropriate patients in an electronic health record. All criteria were converted into logical conditions and the query format permitted the filtering according to the principal diagnosis with inclusion and exclusion constraints.

V. RESULTS AND OUTPUTS

The results produced using the zero-shot prompt engineering strategy showed the possibility of applying a Large Language Model (LLM) to autonomously design clinical trials. The model was also effective in generating a final Phase II protocol for a hypothetical KRAS G12C inhibitor, NewDrugX, that will treat metastatic colorectal cancer. The protocol started with a title of the protocol then continued with a background section describing in details why the KRAS G12C mutation was targeted. This mutation was identified by the model as one of the main causing factors in a mutation in a proportion of colorectal cancers and therefore necessitated the intervention to treat them [7].

The LLM also created a structured trial design section, detailing a single-arm, open-label, multicentre Phase II trial, where the patients would get NewDrugX until disease progression or intolerable toxicity [8]. The model defined inclusion criteria with accuracy and clinical relevancy, as participants had histologically confirmed metastatic colorectal adenocarcinoma, recorded the presence of KRAS G12C mutation, ECOG performance status 0 or 1, and measurable disease.

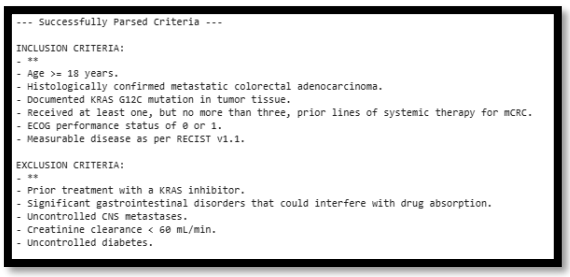
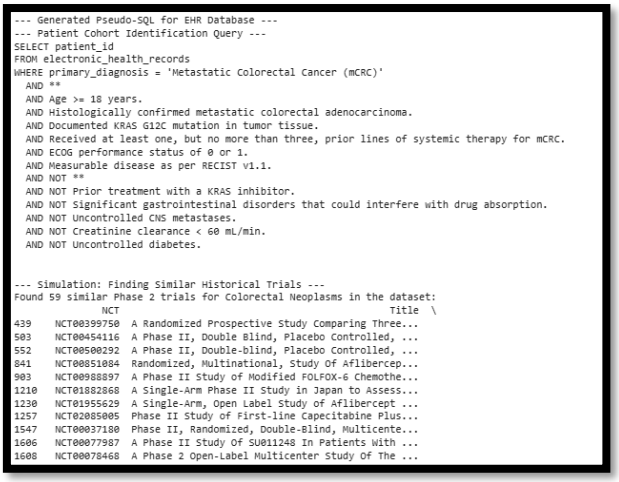


Fig 3: LLM-Generated Clinical Trial Protocol: Section Breakdown

Inclusion and exclusion criteria were programmatically extracted using regular expressions within the text of the protocol. These were then adapted into a pseudo-SQL query to identify an appropriate collection of patients in an electronic health record system [9]. The built query incorporated a filter on primary diagnosis, as well as logical conditions corresponding to the stated inclusion and exclusion criteria, effectively simulating the way in which a trial eligibility algorithm could actually perform.

The second part of the research was the use of the same dataset (Sanofi-funded one) in order to find comparable Phase II colorectal cancer trials. Phase III designs were more abundant; however, some relevant Phase II trials

could be collected, providing a comparative rationale in terms of enrolment number and design considerations [10]. These illustrations confirmed the validity of the output of the LLM by matching historical standards.



```
--- Generated Pseudo-SQL for EHR Database ---
--- Patient Cohort Identification Query ---
SELECT patient_id
FROM electronic_health_records
WHERE primary_diagnosis = 'Metastatic Colorectal Cancer (mCRC)'
  AND **
  AND Age >= 18 years.
  AND Histologically confirmed metastatic colorectal adenocarcinoma.
  AND Documented KRAS G12C mutation in tumor tissue.
  AND Received at least one, but no more than three, prior lines of systemic therapy for mCRC.
  AND ECOG performance status of 0 or 1.
  AND Measurable disease as per RECIST v1.1.
  AND NOT **
  AND NOT Prior treatment with a KRAS inhibitor.
  AND NOT Significant gastrointestinal disorders that could interfere with drug absorption.
  AND NOT Uncontrolled CNS metastases.
  AND NOT Creatinine clearance < 60 mL/min.
  AND NOT Uncontrolled diabetes.


--- Simulation: Finding Similar Historical Trials ---
Found 59 similar Phase 2 trials for Colorectal Neoplasms in the dataset:
                NCT                                              Title  \
439    NCT00399750  A Randomized Prospective Study Comparing Three...
503    NCT00454116  A Phase II, Double Blind, Placebo Controlled, ...
552    NCT00500292  A Phase II, Double-blind, Placebo Controlled, ...
841    NCT00851084  Randomized, Multinational, Study Of Aflibercep...
903    NCT00988897  A Phase II Study of Modified FOLFOX-6 Chemothe...
1210   NCT01882868  A Single-Arm Phase II Study in Japan to Assess...
1230   NCT01955629  A Single-Arm, Open Label Study of Aflibercept ...
1257   NCT02085005  Phase II Study of First-line Capecitabine Plus...
1547   NCT00037180  Phase II, Randomized, Double-Blind, Multicente...
1606   NCT00077987  A Phase II Study Of SU011248 In Patients With ...
1608   NCT00078468  A Phase 2 Open-Label Multicenter Study Of The ...
```

Fig 4: Pseudo-SQL Query for Simulated Patient Cohort Identification

Lastly, a draft summary that could be incorporated into a regulatory filing, like an Investigational New Drug (IND) application, was generated. This overview contained a title of the protocol, rationale, key objectives, and study design, and intended population of patients [11]. The section reflected the informative and formal tone and content arrangement required in regulatory submissions, reinforcing the model as the ability to reproduce industry-standard texts in a zero-shot scenario.

VI. DISCUSSION

In the current study, the outputs generated using the Large Language Model were highly coherent, technically accurate, and contextually coming of age, which proves the feasibility of zero-shot clinical trial design in practice. The protocol that has been produced adhered to the standard rules of structure implemented in professional medical documentation, such as well-defined objectives, logical inclusion and exclusion criteria, and well thought-through trial design [12]. The model could create a lifelike and self-consistent treatment plan out of sparse though meticulously selected historical facts, indicating its capacity to aid protocol development in the early phases of research with little human intervention.

One of the main observations was the sensitivity of the model to timely quality and contextual anchoring. Applying historical trial examples in writing boosted the applicability and concreteness of the produced text significantly. In the absence of such examples, the LLM was inclined to generalise or lapse into template-like answers [13], which points to a significant reliance on informative prompts. This indicates the need to carefully engineer prompts and the significance of basing the model on representative clinical data to produce high-quality results.

Nevertheless, there were also limitations. Although the output was fluent, the model was sometimes hallucinatory--producing semantically thin but plausibly sound statements that are difficult to disprove or statements that were too generic. This brings potential threats with content without stringent expert validation in case such content is employed [4]. Moreover, the factual basis of the model is still subject to the breadth and precision of its training data, which are not always up to date with the latest scientific findings and regulatory changes.

VII. LIMITATIONS

The data was analyzed in a limited pool of Sanofi-sponsored trials, most of which were Phase III, which is a crucial limitation as the study addresses designing Phase II protocols. The model had limited experience with relevant comparator designs due to the few examples at the early phase, which may challenge the generalisability of the output. The dataset can also be affected by representation bias, in that any one condition, geography or demographic could be underrepresented [7]. This threatens the soundness of zero-shot generation, resulting in biased suggestions that do not consider the needs of different people. The method of arming upon historical trials also supposes good in itself no doubt, the exactness and fullness of cause recorded. Moreover, the Large Language Model can produce structured output but is subject to prompt formulation and has no tools to assess scientific validity or clinical relevance.

VIII. ETHICAL CONSIDERATIONS

AI-generated clinical trial protocols necessitate ethical and regulatory considerations despite their clinical efficacy. The first one is the threat to patient safety due to the potential use of AI-generated materials without clinical validation. Hallucinated or unconfirmed information can be taken as truthful because of the commanding tone the model uses. There is also the issue of trial equality because models trained using historically non-inclusive data can end up reinforcing inequalities by ignoring the underrepresented groups [8]. Such protocols, without clear authorship and responsibility may be excluded by regulatory bodies, serving as an obstacle to formal submission. Ethical design has less obvious aspects: risk-benefit analysis, consent, and monitoring that are outside the scope of AI. Hence, although AI can be used to enhance trial design, human oversight will be critical to the ethical integrity, scientific rigour, and patient-centred outcome.

IX. FUTURE WORK

In the future, a few improvements can be made to this approach to enhance its usefulness and safety. Clinical alignment of LLMs can be mitigated by fine-tuning on large sets of verified clinical protocols, which can reduce halls risks. Human-in-the-loop review mechanisms will be incorporated, and outputs will meet ethical, scientific, and regulatory controls before implementation [9]. Moreover, by connecting the model to electronic health record (EHR) systems, it may be possible to conduct dynamic, patient-level simulations, enhancing cohort feasibility assessment. The above developments would fill the gap between generative AI and real clinical research, step towards real adaptive, data-informed trial design.

X. CONCLUSION

This project illustrated how a zero-shot Large Language Model (LLM) can be used to make detailed clinical trial protocols using prior data. Beginning with a selected group of Sanofi-funded clinical trials, studies were sifted, metadata acquired, and prompts designed drawing on typical case studies. The LLM countered with an elaborate Phase II plan of a hypothetical KRAS G12C inhibitor against metastatic colorectal cancer, including critical aspects like purpose, design, eligibility criteria, and dosing. Parsing of inclusion and exclusion criteria were also done to allow simulated patient cohort selection through a pseudo-SQL query. Also, a regulatory summary, which can be included in an Investigational New Drug (IND) application, was created. Its disruptive potential in biomedical innovation is reflected in the recently successful zero-shot approach. It can be key to saving time, education needed to develop trial design and can be used to develop custom protocols, especially in initial stages of trials where resources are limited. The approach can be applied to therapeutic areas of different focus and magnitude based on long-term patterns and well-organized clinical terminology.

XI. REFERENCES

[1] Shah-Mohammadi, F. and Finkelstein, J., 2024, January. Large Language Model-Based Architecture for Automatic Outcome Data Extraction to Support Meta-Analysis. In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0079-0085). IEEE.

[2] Liu, J., Nguyen, A., Capurro, D. and Verspoor, K., 2025. Comparing Text-Based Clinical Risk Prediction in Critical Care: A Note-Specific Hierarchical Network and Large Language Models. *IEEE Journal of Biomedical and Health Informatics*.

[3] Kim, E., Shrestha, M., Foty, R., DeLay, T. and Seyfert-Margolis, V., 2024, December. Structured Extraction of Real World Medical Knowledge using LLMs for Summarization and Search. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 3421-3430). IEEE.

[4] Papanastasiou, G., Dikaios, N., Huang, J., Wang, C. and Yang, G., 2023. Is attention all you need in medical image analysis? A review. *IEEE Journal of Biomedical and Health Informatics*, *28*(3), pp.1398-1411.

[5] De Santis, E., Martino, A., Ronci, F. and Rizzi, A., 2024. From bag-of-words to transformers: A comparative study for text classification in healthcare discussions in social media. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

[6] Gutiérrez, J.D., Rodriguez-Echeverria, R., Delgado, E., Rodrigo, M.Á.S. and Sánchez-Figueroa, F., 2024. No more training: SAM's zero-shot transfer capabilities for cost-efficient medical image segmentation. *IEEE Access*, *12*, pp.24205-24216.

[7] Zhu, T., Afentakis, I., Li, K., Armiger, R., Hill, N., Oliver, N. and Georgiou, P., 2024. Multi-horizon glucose prediction across populations with deep domain generalization. *IEEE Journal of Biomedical and Health Informatics*.

[8] Li, N., Zhou, C., Gao, Y., Chen, H., Zhang, Z., Kuang, B. and Fu, A., 2025. Machine unlearning: Taxonomy, metrics, applications, challenges, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.

[9] Du, R.F., Carbonell, E.L., Huang, J., Liu, S., Wang, X., Shen, D. and Ke, J., 2025. Ethics of Foundation Models in Computational Pathology: Overview of Contemporary Issues and Future Implications. *IEEE Transactions on Medical Imaging*.

[10] De Silva, D., Kaynak, O., El-Ayoubi, M., Mills, N., Alahakoon, D. and Manic, M., 2024. Opportunities and challenges of generative artificial intelligence: Research, education, industry engagement, and social impact. *IEEE Industrial Electronics Magazine*.

[11] Kulsoom, U., Glavin, F.G. and Bendechache, M., 2025. Natural Language Processing and Machine Learning for Analysis of Radiology Reports-A Systematic Review. *IEEE Access*.

[12] Zhang, H., Niyato, D., Zhang, W., Zhao, C., Du, H., Jamalipour, A., Sun, S. and Pei, Y., 2024. The role of generative artificial intelligence in internet of electric vehicles. *IEEE Internet of Things Journal*.

[13] Khalili, E., Priego-Torres, B., Leon-Jimenez, A. and Sanchez-Morillo, D., 2024. Automatic lung segmentation in chest X-ray images using SAM with prompts from YOLO. *IEEE Access*.