Vidhi Patel, Cindy Gao , Shuangrui Chen, Yongzhuo Chen

SPECIFIC AIMS

With healthcare digitization, Electronic Health Records (EHRs) have grown exponentially, offering improved patient care. However, the challenge lies in effectively managing vast patient data for clinical decision-making. This project aims to employ artificial intelligence (AI) and natural language processing to navigate EHR complexities, focusing on optimizing cancer diagnosis through Large Language Models (LLMs).

While previous studies showcased GPT-3 and -4's diagnostic accuracy, this study innovates by optimizing LLMs for oncology diagnosis through a neural network-based question embedding strategy. Unlike past research, it addresses limitations of using a pre-existing LLM, particularly the lack of optimization for the specific diagnosis task.

Initial robustness assessment on LLM-generated questions from the Handbook of Oncology has revealed promising evidence in its ability to accurately identify embeddings related to the prompt in question. Alinger, a neural network, will then be used to effectively map the random forest generated positive embeddings to the most relevant section within the document space, ultimately, improving the LLM’s precision and ability in locating precise information. Based on this, we hypothesize that an LLM can be optimized for oncology diagnosis through a neural network based question embedding strategy.

These hypotheses will be addressed in the experiments of the following Specific Aim:

Specific Aim 1: Enhance LLM Guidance For Question-Answering: Despite promising results, the LLM exhibited inconsistencies in selecting pertinent handbook chunks, especially in broader searches. Retraining or fine-tuning is challenging due to resource constraints. Integrating the neural network Aligner will resolve inconsistencies in LLM's selection of relevant handbook chunks for oncology question-answering.

We anticipate that incorporation of the proposed Aligner model will aid in chunk retrieval for a given question and thus increase LLM performance for medical case diagnosis specific to oncology. Furthermore, identification of relevant chunks within the document can be incorporated as reference during question-answering, helping the user efficiently refer to the larger corpus if necessary. If so, this could prove to be a promising contribution in assessing potential LLM incorporation in clinical practice.

Vidhi Patel, Cindy Gao , Shuangrui Chen, Yongzhuo Chen

RESEARCH STRATEGY

SIGNIFICANCE: Nowadays, with the advancements of technology, it has steadily propelled the healthcare industry into the digital era. The transition to storing patient information in Electronic Health Records (EHRs) has resulted in a significant increase in volume of patient data. While this digital transformation offers enhanced patient care and streamlined healthcare processes, it also presents significant challenges in effectively managing, interpreting, and optimally leveraging the "usefulness" of this vast repository of information for clinical decision-making.

Recognizing these challenges, this project aims to harness the power of artificial intelligence (AI) and natural language processing to navigate the complexities of EHRs, which includes but are not limited to Electronic Medical Records (EMRs), lab results, and clinical notes.

Specifically, in this study, we aim to leverage Large Language Models (LLM) to help optimize the classification and detection of cancer in patients in reference to the comprehensive knowledge from the Handbook of Oncology. Through trying to improve the accuracy and relevance of the LLM response through text embeddings and pattern findings, this could potentially pave the way to higher precision and more personalized medication, thus essentially optimizing patient care in oncology.

INNOVATION: Previous studies have demonstrated GPT-3 and -4's ability to accurately diagnose clinical cases [1,2,3,4]. Additionally, there has also been research on the performance of GPT on cancer diagnosis. However, these studies had considerably low training datasets and did not seek to optimize the LLM based on the diagnosis task at hand. Other studies focused on oncology specific information retrieval using transformer based-LLMs show promising results for information retrieval within specialized medical fields. One limitation of these studies is that they focus on basic multiple-choice questions rather than open-ended formats [1,2]. This study is innovative in the terms of its approach on LLM utilization as it optimizes the model inputs through neural network generated question embeddings for open-ended questions. The proposed study has the potential to improve cancer diagnosis and subsequently patient outcomes by leveraging generative AI as an augmented clinical partner as well as revealing potential areas of strength as well as shortcomings of LLM integration as support in clinical practice.

RESEARCH PLAN: We propose that an LLM can be optimized for oncology diagnosis through a neural network based question embedding strategy.

LLM Robustness: During the initial phase of this study, the goal is to assess "how well" an LLM can parse the given text, in this case, the Handbook of Oncology. To do this, a simple experiment was conducted where the LLM was tasked with generating questions given a specific chunk from the Handbook of Oncology.

Random Forest: To assess the questions posed by the LLM, a Random Forest Algorithm was implemented. The input of the Random Forest algorithm was the text embedding of the questions posed by the LLM, with a binary outcome of either 1 or 0; if the question was answerable from the Handbook of Oncology, or not, respectively. The results were promising, with the Out-of-Bag (OOB) Area Under the Curve (AUC) remarkably high at 0.9983, and the Precision-Recall Area Under the Curve (PR-AUC) at 0.9986.

Aligner: Alinger, a neural network, will be used as a function mapping the question space to the document space, to help identify the most related chunk of the document to the question. Given the results from the Random Forest algorithm, if the question posed by the LLM was indeed related to the Handbook of Oncology, the embedded question would be used as the input for the Alinger neural network to generate an output embedding, which is a given chunk from the document space (Handbook of Oncology). The distance between the output embedding and the document space embeddings belonging to the Handbook of Oncology will be calculated; the goal is the find the smallest distance between the output and different embeddings, which in translation gives us the chunk that is the most related to the question posed by the LLM. Thus, the LLM can seek reference or answer the posed questions from the embedding chunk with the lowest distance.

SPECIFIC AIM 1: Enhance LLM Guidance for Question-Answering

Hypothesis: The integration of a neural network will help address issues in oncology specific question answering abilities of LLM when using the Handbook of Oncology.

Rationale: While the initial experiment demonstrated promising results, the LLM exhibits inconsistency in selecting the most pertinent handbook chunk, particularly when expanding the search to broader selections (e.g., top 10 or top 100 closest chunks). This inconsistency poses a challenge by leading to inaccuracies and the presentation of unrelated information in LLM responses. While a potential solution to this issue could include retraining the LLM, this would take lots of time and money. Another potential solution is fine-tuning, yet we lack the labeled data and it's difficult to assess the model's performance in this specific domain.

Experimental Approach:

- Document Retrieval Optimization:
    - Analyze LLM performance by investigating the ability to identify the closest, top 10, and top 100 most relevant chunks for each query. Performance will be assessed using mean accuracy for each selection category (e.g., top 10, top 100).
    - Address challenges faced in achieving high accuracy during document retrieval, ensuring a more precise selection of handbook sections for responding to queries.
        - Build a neural network - ALIGNER - to modify embedding of questions so the chunks of the document are more accurately matched with the question. A preliminary question bank consisting of 1033 questions, spanning over 500 chunks of text, will be used to optimize question embeddings.

Interpretation of Results: Enhanced relevance detection, especially in broader search selections, will help mitigate inconsistencies and inaccuracies, leading to more precise and informative responses in the context of oncology diagnostics.

Potential Problems and Alternative Approaches: Alternative approaches may involve exploring different machine learning algorithms or leveraging additional features for relevance detection.

Citations

1. Benary M, Wang XD, Schmidt M, et al. Leveraging Large Language Models for Decision Support in Personalized Oncology. *JAMA Netw Open*. 2023;6(11):e2343689. Published 2023 Nov 1. doi:10.1001/jamanetworkopen.2023.43689

2. Blum J, Menta AK, Zhao X, Yang VB, Gouda MA, Subbiah V. Pearls and pitfalls of ChatGPT in medical oncology. Trends Cancer. 2023;9(10):788-790. doi:10.1016/j.trecan.2023.06.007

3. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J Med Syst*. 2023;47(1):33. Published 2023 Mar 4. doi:10.1007/s10916-023-01925-4

4. Eriksen, A. V., Möller, S., & Ryg, J. (2023). Use of GPT-4 to diagnose complex clinical cases. NEJM AI, 1(1), AIp2300031.

5. Kanjee Z, Crowe B, Rodman A. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. *JAMA*. 2023;330(1):78-80. doi:10.1001/jama.2023.8288