Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240521

Unlocking the Potential of Free Text in Electronic Health Records with Large Language Models (LLM): Enhancing Patient Safety and Consultation Interactions

Pushpa KUMARAPELI [a,1], Tia HADDAD[a] and Simon de LUSIGNAN[b]
[a] *School of Computer Science and Mathematics, Kingston University London, UK*
[b]*Nuffield Department of Primary Care Health Sciences, University of Oxford, UK*
ORCiD ID:Pushpa Kumarapeli https://orcid.org/0000-0002-9552-6835
Tia Haddad https://orcid.org/0009-0009-5187-1168
Simon De Lusignan https://orcid.org/0000-0002-8553-2641

Abstract. Computer-mediated clinical consultation, involving clinicians, electronic health record (EHR) systems, and patients, yield rich narrative data. Despite advancements in Natural Language Processing (NLP), these narratives remain underutilised. Free text recording in EHRs allows expressivity, complements structured data from clinical coding systems, and facilitates collaborative care. Large language models (LLMs) excel in understanding and generating natural language, enabling complex dialogue processing. Integrating LLM tools into consultations could harness the untapped potential of free text to identify patient safety concerns, support diagnosis and provide content to enhance clinical-patient interactions. Tailoring LLMs for specific consultation tasks through pre-training and fine-tuning is viable. This paper outlines approaches for adopting LLMs in primary care and suggests that using fine-tuned LLMs with prompt engineering could enhance computer-mediated clinical consultation cost-effectively.

Keywords. Primary care, Patient safety, Domain specific LLM

1. Introduction

Computer-mediated clinical consultations generally involve a triadic relationship between the patient, clinician, and the electronic health record (EHR) system. In a previous study we analysed triadic consultation task distribution and its effects on data recording and social interactions [1]. Reviewing past medical encounters and patient history, along with data entry, both structured and in free text, constitute a significant portion of clinical consultations. Data collected during clinical consultations is crucial source for epidemiological research, care planning, policymaking, and generating evidence-based practices. However, free text entry during consultations is primarily validated for administrative and memory-aid purposes rather than those secondary uses.

[1] Corresponding Author: Pushpa Kumarapeli; E-mail: P.Kumarapeli@kingston.ac.uk.

The primary challenge in interpreting free-text content in EHRs lies in the technical aspects, specifically in automating this process and ensuring anonymisation. Despite advancements in natural language processing (NLP) technologies, the full potential of the narratives captured into EHR in these encounters remains largely unexplored. Large language models (LLMs) show promise in understanding and generating natural language, with potential for enhancing consultation interactions.

We conducted a narrative review of journal articles to establish domain specific LLM adoption approaches, then narrowed our focus to case studies in the healthcare domain. Our critical analysis and interpretation of the findings centred on how well these LLM adoption approaches meet the specific demands of primary care consultations and their potential to enhance consultation tasks and outcomes.

2. The Role of Free-text in the Clinician-EHR-Patient Consultation Interactions

In the primary care settings, where general practitioners serve as gatekeepers, doctor-patient interactions are fundamental. Clinical coding systems utilised in EHRs play a crucial role in clinician-EHR interactions. However, coding processes can be time-consuming, and there exists a diversity of usage patterns [2]. Decision-making in this context is further influenced by professional ethics, obligations, and the dynamics of doctor-patient relationships, which consequently impact the use of EHR.

Free text offers expressivity, allowing clinicians to adopt their own story-telling structures and can be used as a strategy to accommodate unexpected outcomes [3]. The collaborative nature of patient care delivery is supported by free text, as records can be maintained to easily reconstruct the patient's story later, either by the authors themselves or other members of the healthcare team.

While coded segments within EHRs are typically prioritised for routine data extraction and analysis, the free text component serves multiple purposes. It acts as a space for thought, decision rationalisation, and recording, enhancing clarity for coded entries. The free text often includes qualifiers, negations, and contextual details, capturing nuances that may be missed by codes. Relying solely on coded data may result in overlooked diagnoses documented only in free text, potentially causing delays in diagnosis. Studies on full-text reviews are limited due to anonymisation costs.

3. Leveraging the Developments in LLMs to Support Consultation Tasks

Advancements in NLP have facilitated the development of complex dialogue processing capabilities through contextual models and Transformer-based language models [4]. These models interpret the context and intent of text using either a probabilistic approach or relying on neural networks (NNs). LLMs are a subtype of deep learning (DL) models focusing on NLP, trained on significantly larger amounts of data and parameters compared to their regular counterparts, enhancing their performance in more complex knowledge domains. Pretraining these LLMs involves exposing them to vast corpora of text. Notable examples of foundation models include those built on the Transformer architecture, such as OpenAI’s GPT-n (Generative Pre-trained Transformers) series and Bidirectional Encoder Representations from Transformers (BERT). These models demonstrate the ability to understand context, semantics, and a wide range of language structures. The use of scalable self-supervised learning techniques, enable them to adapt effectively to various application scenarios. These models excel in their ability to generalise, adapt, and self-supervise, leveraging extensive pre-training.

After reviewing current research in NLP, specifically the LLM sub-domain, and considering the reported limitations of EHR in supporting computer-mediated consultation tasks, we identified three key areas where EHR integrated LLM layer could offer significant benefits (figure 1); (1) differential diagnosis: recognising potential or incorrect diagnoses, (2) safety alerting: comprehending safety risks and hazards (3) generative consultation support resources: LLM-generated content to enhance consultation interactions and outcomes. Given the necessity for tailored LLM integration, which must accommodate domain expertise and the complexities of triadic interactions, we proceeded to investigate potential methods for achieving this goal.

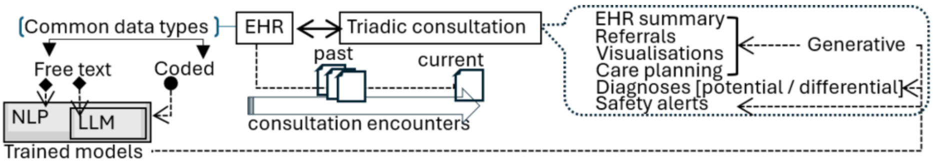

Figure 1. Potential for NLP-LLM elements to enhance consultations through alerts and generative content

4. Towards Specialised LLMs to support consultation tasks and outcomes

Integrating NLP tools to support consultation interactions can benefit from LMs tailored to produce task- and domain-specific outcomes. These can be achieved through LM training or prompt engineering (PE) [5]. LM training typically involves pre-training, fine-tuning, and alignment stages (figure 2). Pre-training allows LMs to absorb knowledge from unlabelled text data. Fine-tuning ensures model adherence to instructions, while alignment focuses on providing useful user responses. LMs perform better with contextual prompts during pre-training and fine-tuning. PE involves crafting prompts to enhance LM relevance and quality, offering a cost-effective alternative for targeted performance improvements.

4.1. Training Phase LLM Modifications

Fine-tuning involves enhancing Transfer learning by adjusting the weights of the model with new data. Fine-tuning can be also achieved by augmenting the model using Adapters, utilising a subset of original parameters, and adding new layers. Supervised fine-tuning uses a smaller dataset with input-output pairs, to refine the model's performance based on specific instructions. Fine-tuning existing LMs is cost-effective compared to training from scratch. Given the abundance of healthcare data and evidence-based practices, LLMs likely already have domain knowledge. For smaller models, selective fine-tuning, like Parameter Efficient Fine Tuning (PEFT) targets specific layers.

In the alignment stage, Reinforcement Learning with Human Feedback (RLHF) approaches utilise reward models to rank and refine outputs. This often leads to outputs that are better aligned with human values, exhibit improved adaptability, and ensure safety, while also reducing ambiguity. However, scalability issues, resource intensiveness, and the potential for bias are notable disadvantages. RLHF alternatives such as Hindsight Instruction Relabelling (HIR), Constitutional AI, Direct Preference

Optimisation (DPO), Reinforced Self-Training (ReST), and Reinforcement Learning with AI Feedback (RLAIF) address these weaknesses by leveraging high-quality human-defined rules, employing self-supervised learning, conducting task-specific simulations, or utilising targeted datasets.

4.2. Prompt Engineering

This approach involves modifying the prompts given to the model to steer its outputs towards desired task-specific outcomes [6]. The existing PE guidelines for the healthcare domain are limited. However, the PE guidelines established at a general level provide a foundational basis for exploring opportunities to extend them. Some of the commonly employed PE strategies include providing examples with context, providing context before questions, supplying context tokens, adding context to prompt, and conditioning the prompt with specific identities (figure 2). PE through templated inputs can be a more pragmatic adoption for the healthcare settings. Instruction tuning enables the model to infer the input based on predefined instructions.

5. Challenges and Opportunities

Research has demonstrated that LLMs can perform at a level comparable to humans when responding to medical questions [7]. However, their capacity to generate natural language responses encompassing complex biopsychosocial elements, and maintaining relevance to medical ontologies, coding, and classification systems, requires further refinement [8]. The mental workload, decision-making involved in disease-illness distinctions, and the consideration of the patient as a biomedical entity, sentient being or an autonomous individual based on the behavioural aspects, add significant complexities medical practice. Nevertheless, with a systematic approach, training, and fine-tuning, LLMs can be enhanced to address these complexities more effectively. The existence of clinical coding and classification systems, standardised vocabularies such as UMLS, and incentives encouraging coding practices present opportunities that position the healthcare domain uniquely.

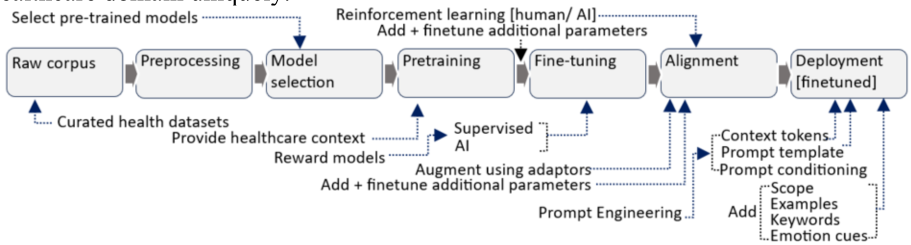

Figure 2. Possible approaches for LLM customisation for primary care

LLM incorporation in consultations could vary based on clinical nuances. LLMs fine-tuned on clinical notes have shown high accuracy when establishing symptoms critical for treatment outcomes [7]. Combining prompt engineering with PEFT to interpret Hospital Discharge Summary papers has produced accurate summaries of complex medical narratives [8]. When decision aids possess good usability characteristics and can be seamlessly integrated into routine practice, clinicians recognise their value. This integration facilitates more meaningful patient-oriented discussions and supports patient safety aspects holistically by providing useful reminders and alerts during decision-making processes. However, one barrier to LLM adoption in healthcare is the lack of clarity surrounding how data is collected by providers. This raise concerns related to Information Governance (IG) and confidentiality. Additionally, the drawbacks of LLM integration include disparities caused by biased outcomes, challenges related to EHR integration and quality control, liability issues, explainability concerns, and ethical considerations. Overall, given the challenges and opportunities, it is important to recognise the potential value of adopting LLMs to enhance the clinician-patient interactions of the consultations. There is a need to recognise and explore various approaches to LLM specialisation further.

6. Conclusions

While LLMs demonstrate capabilities in addressing medical queries, their capacity to produce nuanced natural language responses in clinical settings requires further refinement. Challenges persist in integrating complex biopsychosocial elements and aligning with medical ontologies and coding systems. This study suggests a targeted approach to training LLMs specifically for primary care, incorporating interventions during both pre-training and fine-tuning stages. These interventions, combined with prompt engineering methods, have the potential to offer cost-effective and scalable LLM solutions for point-of-care settings. Addressing concerns about data collection transparency is essential for broader LLM adoption. LLMs can play role in enhancing patient safety, supporting diagnosis, and dynamically creating consultation resources to facilitate interactions. This can enhance consultation tasks, outcomes, and support, while also aiding epidemiological research and care management. Continued research and development are essential for fully realising the potential of LLMs, particularly in computer-mediated consultations, where their impact can be most significant.

References

[1] Kumarapeli P, de Lusignan S. Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study. Journal of the American Medical Informatics Association. 2013 Jun 1;20(e1):e67-75. doi: 10.1136/amiajnl-2012-001081
[2] Zghebi SS, Reeves D, Grigoroglou C, McMillan B, Ashcroft DM, Parisi R, Kontopantelis E. Clinical code usage in UK general practice: a cohort study exploring 18 conditions over 14 years. BMJ open. 2022 Jul 1;12(7):e051456. doi: 10.1136/bmjopen-2021-051456
[3] Millares P. Consultation analysis: use of free text versus coded text. Health and Technology. 2021 Mar;11(2):349-57
[4] Liu Z, Zhong A, Li Y, Yang L, Ju C, Wu Z, Ma C, Shu P, Chen C, Kim S, Dai H. Tailoring large language models to radiology: A preliminary approach to llm adaptation for a highly specialized domain. In International Workshop on Machine Learning in Medical Imaging 2023 Oct 8 (pp. 464-473). Cham: Springer Nature Switzerland.
[5] Gabriel I. Artificial intelligence, values, and alignment. Minds and machines. 2020 Sep;30(3):411-37.
[6] Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. Journal of Medical Internet Research. 2023 Oct 4;25:e50638. doi: 10.2196/50638
[7] Chiang CC, Luo M, Dumkrieger G, Trivedi S, Chen YC, Chao CJ, Schwedt TJ, Sarker A, Banerjee I. A large language model‐based generative natural language processing framework fine‐tuned on clinical notes accurately extracts headache frequency from electronic health records. Headache: The Journal of Head and Face Pain. 2024 Apr;64(4):400-9.
[8] Goswami J, Prajapati KK, Saha A, Saha AK. Parameter-efficient fine-tuning large language model approach for hospital discharge paper summarization. Applied Soft Computing. 2024 May 1;157:111531.