LLM-Generated Clinical Histories: Evaluating Readability and Efficiency for Radiologic Interpretation

Shawn Lyo, MD, Neuroradiologist and Imaging Informaticist, University of Pennsylvania/Baptist Health South Florida
Satvik Tripathi; Ali Tejani, MD; David Vu, MD; Siddhant Dogra, MD; Hamza Alizai, MD; Tessa Cook, MD, PhD, CIIP, FSIIM, FCPP

Introduction

Radiologists face significant time pressure when interpreting imaging studies, especially during on-call shifts. Reviewing lengthy clinical documentation is time-consuming, with critical information often buried within extensive clinical notes. Large language models (LLMs) have demonstrated capabilities in synthesizing and summarizing medical text, suggesting potential utility in automatically generating focused, relevant clinical histories. However, the ability of LLMs to produce clinically useful and accurate history summaries for radiology interpretation has not been systematically evaluated.

Hypothesis

LLM-generated clinical histories will be more readable and efficient to review than manually extracted clinical documentation while maintaining clinical accuracy.

Methods

A two-stage summarization process was developed using an institutionally approved Azure OpenAI GPT-4o model utilizing ensemble prompting and universal self-consistency techniques. For each of 30 imaging studies, concise clinical histories were generated by synthesizing information from imaging orders, recent clinical notes, and prior imaging reports. Stage 1 generated a detailed timeline-based summary, while Stage 2 produced a single-paragraph condensed summary (Figure 1). Text length, readability scores (Flesch-Kincaid, SMOG, Coleman-Liau), and clinical completeness were analyzed.

Results

The generated clinical histories significantly reduced text while maintaining clinical utility. Source texts averaged 2,021 words. Stage 1 and 2 summaries reduced text length by 82.6% and 93.8% respectively. Increasing levels on readability indices suggest that requisite technical complexity for interpretation was preserved (Figure 2). Preliminary evaluations suggest that essential medical information was preserved.

Conclusion

LLM-based clinical history generation approach effectively reduces text length by over 93% while retaining essential clinical information. This has the potential to enhance radiologist workflow efficiency by minimizing the time spent reviewing histories and ensuring that critical details are accessible rather than buried in extensive clinical documentation. Further studies will evaluate its direct impact on time savings and diagnostic accuracy.

Figure(s)

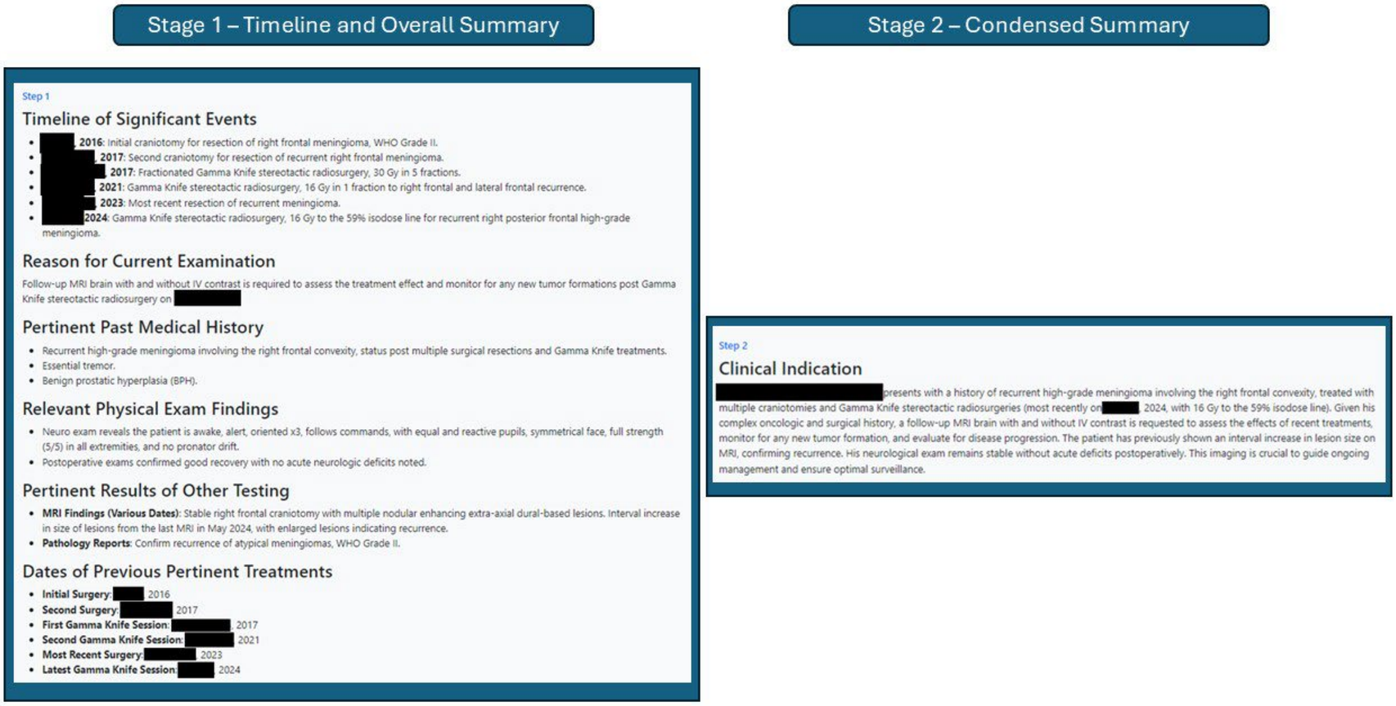

Figure 1. Examples of automated clinical history generation outputs. Stage 1 demonstrates a detailed summary that includes a timeline of significant clinical events, pertinent medical history, physical exam findings, and previous treatments. Stage 2 condenses this information into a single-paragraph summary, providing a concise yet comprehensive clinical indication.

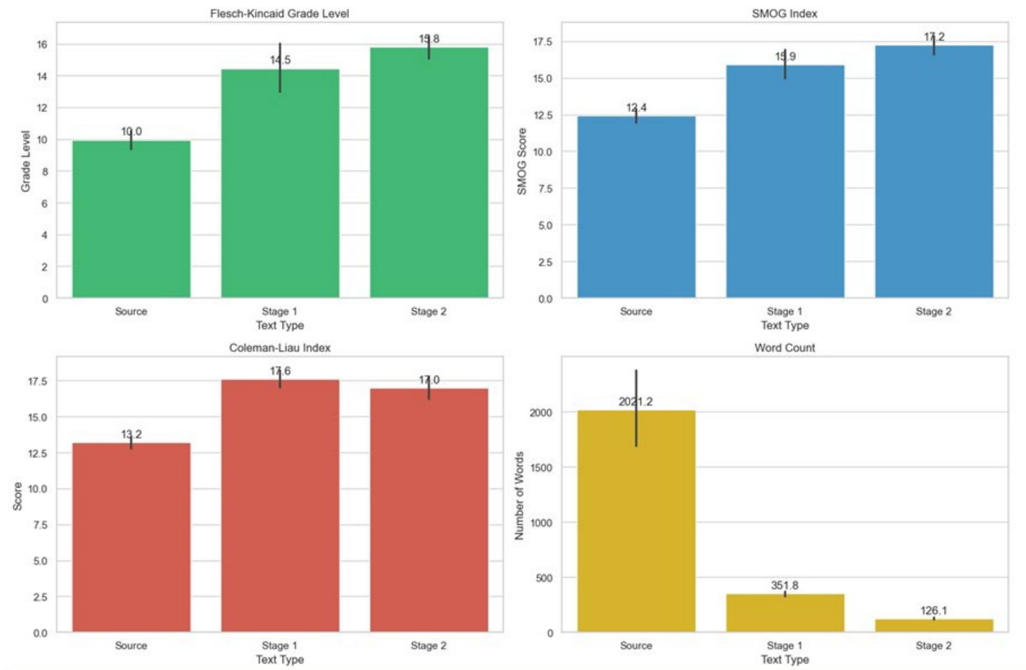

Figure 2. Comparison of readability indices and word count between original clinical data (Source), timeline and summary of pertinent information (Stage 1), and condensed summary (Stage 2). Generated histories were significantly shorter than the source clinical data but featured a higher proportion of complex words and sentence structures, as reflected in the increased Flesch-Kincaid Grade Level, SMOG Index, and Coleman-Liau Index.

Keywords

Applications; Artificial Intelligence/Machine Learning; Clinical Workflow & Productivity; Emerging Technologies; Provider Experience; Quality Improvement & Quality Assurance