Leveraging Large Language Models to Enhance Radiology Communication in the Emergency Department: A Quality Improvement Project

Author: Anna Rachel Menezes

Background: Clear communication of radiology results to patients in the Emergency Department (ED) is essential for patient understanding, satisfaction, and informed decision-making. However, ED residents often face the dual challenge of explaining them quickly to patients in high-pressure, time-constrained environments. Research indicates that communication gaps can lead to misunderstanding, anxiety, reduced adherence to treatment, and potentially compromised followup care. While structured templates and discharge summaries exist, they are not uniformly applied, and there is no standardized approach to ensuring that radiologic findings are explained in plain, patient-friendly language. Recent developments in artificial intelligence (AI), particularly large language models (LLMs), offer a potential solution. LLMs have demonstrated the ability to generate accurate, readable, and context-sensitive text in multiple domains, including medicine. Objective: This quality improvement project aims to evaluate whether a large language model (Google's PaLM, accessed via Google Colab) can support ED residents by generating patientfriendly explanations of common imaging findings. Our primary goal is to increase the percentage of radiology explanations rated as "highly understandable". We define this as a Flesch-Kincaid Grade Level 6 - 8, with positive qualitative assessment—by at least 10% over a 3-month intervention period. Methods: We designed a two-phase intervention based at a Level 1 trauma center. Using Epic's SlicerDicer tool, we identified the top 10 most frequently encountered imaging-related diagnoses in the ED, such as subdural hematoma and subarachnoid hemorrhage. These findings were used to create standardized prompts simulating real-world clinical communication tasks. In Phase 1 (baseline), ED residents were asked to explain each of the 10 findings to a hypothetical patient. Their responses were assessed for (1) medical accuracy, (2) readability using the Flesch-Kincaid Grade Level, and (3) qualitative engagement and clarity using a structured rubric. We use LLM-as-a-judge which is a state-of-art model used to compare language responses for accuracy, readability, and tone. Simultaneously, the Google PaLM model is prompted with the same scenarios to generate AI-based explanations. These outputs were synthetically generated and did not involve protected health information (PHI). The AI was iteratively refined through a process of prompt engineering based on resident feedback in blinded assessments. Residents were asked to compare AI- and human-generated responses, determine which they would use or prefer in real clinical interactions, and determine which sounded more human. Phase 2 involved implementation and refinement. AI-generated explanations were modified using an iterative prompt engineering loop: Initial Prompt → LLM Output → Resident Feedback → Prompt Revision → New Output. This loop was repeated until the explanations reached a defined threshold of acceptability (≥75% of residents rating the output as "clear and usable"). Results (anticipated): We hypothesize that AI-generated explanations, after refinement, will demonstrate a ≥10% improvement in readability scores (Flesch-Kincaid) compared to baseline resident explanations. We also expect that at least 75% of AI responses will be rated equal or superior in clarity compared to human-generated counterparts. Balancing measures will include resident perceptions of whether the AI added or reduced cognitive burden and the risk of medical oversimplification. Importantly, no AI outputs will be used directly with patients, and the tool is not integrated with Epic or PHI, aligning with current institutional guidelines. Conclusion: This project explores how LLMs can be ethically and effectively used to support resident education and improve the clarity of radiologic communication in emergency settings. By combining real-world diagnostic trends (via SlicerDicer) with iterative prompt optimization, we aim to demonstrate that AI tools can serve as communication co-pilots—not replacements—for physicians. Potential future applications include automated generation of radiology summaries in MyChart, real-time resident support during high-volume shifts, and scalable language adaptation for diverse patient populations. Long-term, such tools could support emotionally intelligent patient care in complex settings, such as serious illness conversations, end-of-life discussions, or trauma recovery.