Enhancing Radiology Resident Education Through Locally Deployed AI/ML-Driven Report Feedback: A Pilot Study to Evaluate Feasibility and Impact

Orhan Unal, PhD, Assistant Professor, University of Wisconsin School of Medicine & Public Health
Taylor Sellers, MD; Erik Winterholler, MD; Tan Nicholas, MD; Isha Pathak, MD; Allison Grayev, MD; John Garrett, PhD

Introduction

Attending radiologist feedback on residents' preliminary reports is vital for radiology education. However, increasing imaging volumes have reduced opportunities for direct feedback, leaving attending revisions as an underutilized learning resource. Extracting structured, actionable feedback from these revisions is challenging due to variability and lack of guidance.

Hypothesis

A locally deployed AI/ML solution utilizing large language models (LLMs) can efficiently analyze and compare preliminary and final radiology reports, providing actionable feedback while safeguarding PHI through local deployment. This solution could serve as a practical educational tool for radiology training.

Methods

Impressions from overnight preliminary reports and corresponding final reports for chest X-ray and head CT studies were extracted via SQL queries and formatted into anonymized CSV tables. Leading open-source LLMs, including LLaMA3.1, Gemma2.5, and Mixtral, were tested with various zero-shot prompts designed to analyze differences between preliminary and final reports. Key aspects included clarity, relevance to training, clinical accuracy, and specificity. Initial subjective assessments evaluated the outputs' alignment with educational goals, refining prompt design to optimize model performance in addressing radiology report nuances.

Results

Our initial findings demonstrate that LLMs effectively identify differences in clarity, training relevance, clinical accuracy, and specificity between preliminary and final reports. Performance varied with prompt structure, emphasizing the importance of prompt engineering in achieving meaningful results. Certain prompts generated outputs more aligned with clinical and educational objectives, confirming the feasibility of using LLMs to produce structured feedback. These results support the potential of locally deployed AI/ML tools to enhance traditional feedback mechanisms while maintaining privacy and data integrity. Quantitative analysis is ongoing to further validate these findings and optimize model outputs.

Conclusion

Locally deployed LLMs demonstrate robust performance in providing structured feedback to radiology residents while ensuring PHI protection. These findings highlight the potential for AI-driven tools to supplement traditional feedback, advancing radiology education and maintaining high standards of privacy and quality.

Keywords

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