2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Large Language Model-Assisted Analysis of Planning Peer Review Data to Drive Quality Improvement in MR-Guided Online Adaptive Radiotherapy

Siqiu Wang, PhD; Christopher Kabat, PhD; Jingying Lin, PhD; Ruiqi Li, PhD; Sean Domal, PhD; Tsuicheng Chiu, PhD; Fan Chi Su, PhD; Ying Zhang, PhD; Justin Visak, PhD; Mu-Han Lin, PhD.
University of Texas Southwestern Medical Center

ABSTRACT

This study evaluates large language model (LLM)-assisted analysis of physics peer-review data from MR-guided online adaptive radiotherapy.

Over one year, 599 plans underwent upstream review using structured checklists and free-text comments. A prompt-engineered LLM workflow, developed with 99 test and 50 validation cases, identified and categorized clinically relevant issues. It detected 798 issues, most commonly involving structures, IMRT constraints, physician intent, and plan quality. Mean review time decreased from 37.6 to 18.5 minutes.

The approach standardizes heterogeneous reviewer feedback, supports targeted quality improvement, improves resource efficiency, and may be transferable to other documented radiation therapy plan-check processes at large scale.

INTRODUCTION

Magnetic resonance imaging-guided online adaptive radiotherapy (MRgRT) requires a complex planning process, during which an upstream peer review prior to physician approval plays an essential role in early error detection and supporting safe and efficient online adaptation. Analyzing the peer review data gives critical insights for process improvement. However, extracting actionable items from the large volume of peer review data from different reviewers can be resource-intensive and difficult to scale. Leveraging a unique real-world structured dataset of physics-led peer review, this study evaluates the use of large language model (LLM)–assisted analysis to support the evolution of peer review processes toward safer and more efficient MRgRT workflows.

METHODS AND MATERIALS

For the past year, an upstream peer review process was implemented prior to physician approval for MRgRT plans, with findings documented using a structured checklist with free-text comments. Over one year, 599 cases underwent review, generating a large pool of clinically annotated data. Using a manually-reviewed subset (99 test cases, 50 validation cases), a LLM prompt-engineering workflow was developed to identify issue-bearing comments, normalize reviewer variability, and classify findings into clinically-relevant categories. Peer review duration and issue distributions were analyzed to assess workflow efficiency and practical applicability.

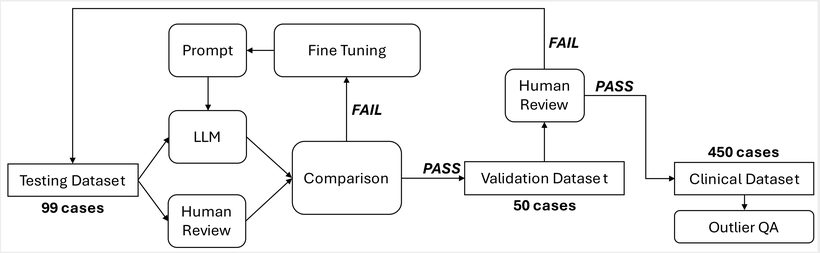

Figure 1: The proposed prompt-engineering workflow to facilitate large-scale analysis of peer review data.

Category	Issue Item
Structures	All required organs-at-risks for the treatment site are sufficiently contoured.
IMRT Constraints	Important DVH criteria have corresponding IMRT constraints/objectives.
Physician Intent	DVH criteria in the treatment planning system matches documented directives.
Plan Quality	Hotspot/max dose distribution is appropriate.
Background	Plan sum is done with prior irradiation.
Plan Parameters	Plan setup and treatment couch indexing matches simulation.
Beams	Beams do not shoot through couch edge, compression, or critical organs.
Schedule	Correct treatment frequency is selected.

Table 1: The most frequently-identified issue in each major category.

RESULTS

LLM-assisted analysis identified 798 issues, most frequently detected in Structures (28.2%), IMRT Constraints (19.3%), Physician Intent (13.9%), and Plan Quality (10.4%). The distribution and prioritization of identified issues were consistent with expectations from experienced reviewers. Over the implementation period, the mean review duration reduced from 37.6 to 18.5 min, reflecting improved efficiency and more focused use of physics resources within the MRgRT workflow.

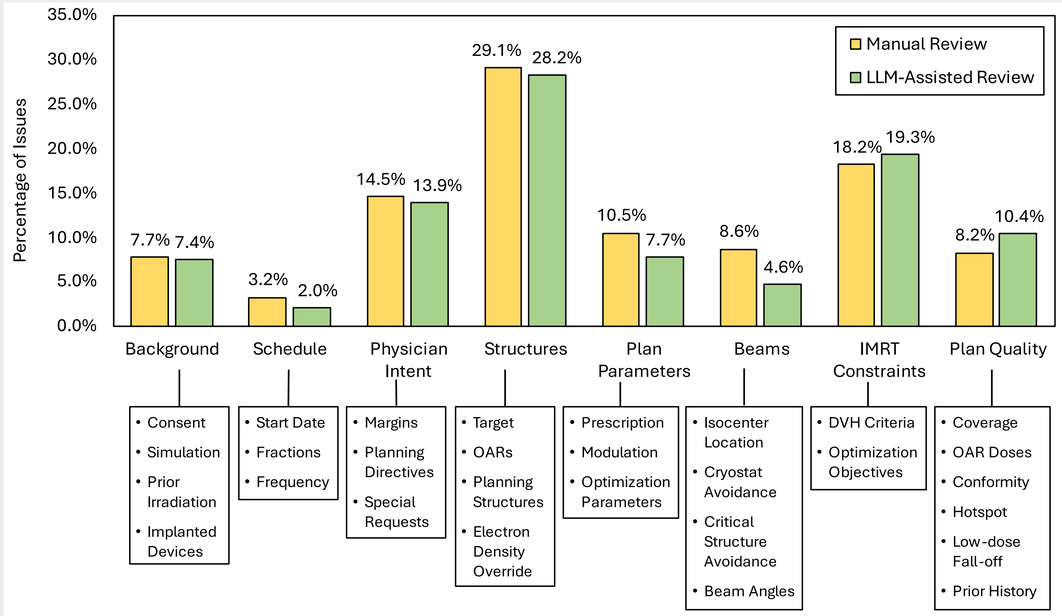

Figure 2: Number of issues identified in each major categories by manual review vs. LLM-assisted analysis of the peer review data for 599 MR-Linac plans over one year of peer review implementation. The largest discrepancies shown in Plan Parameters and Beams can potentially be results of concurrent efforts to standardize planning templates.

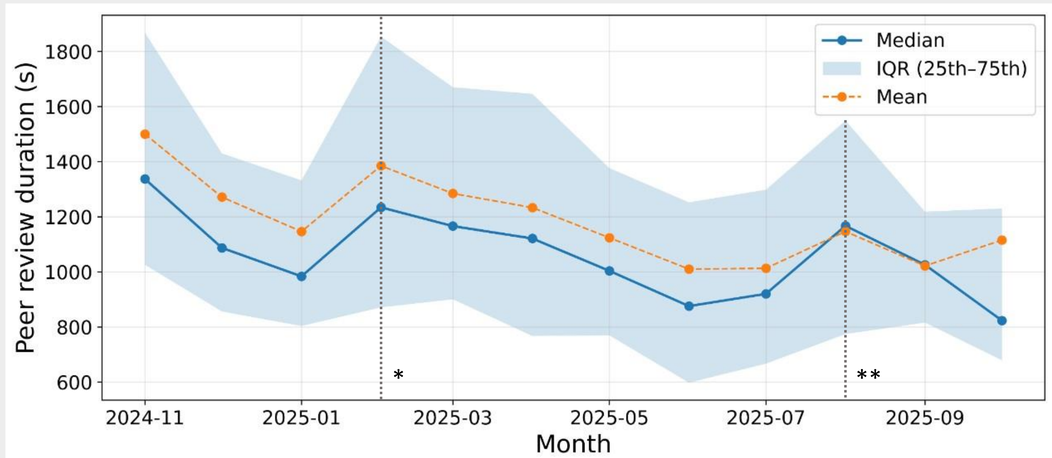

Figure 3: Duration of peer reviews over the 1-year implementation. Surges in duration coincides with *introduction of new checkers and **new motion management tools

CONCLUSIONS

LLM can be effectively leveraged to analyze clinical MRgRT peer review data to inform targeted quality improvement measures. This approach facilitates seamless MRgRT adaptation and provides a pragmatic pathway toward sustainable peer review in the resource-intensive adaptive radiotherapy environments.

CONTACT

Siqiu Wang
The University of Texas Southwestern Medical Center
Dallas, TX
Siqiu.Wang@utsw.edu

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com