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

Automated Prescription and Dose Goal Generation from Free-text Physician Orders using Frontier Large Language Models

Sizhuo Meng[1], Matthew Ladra[1], Curtiland Deville[1], Xun Jia[1], Heng Li[1], William T. Hrinivich[1]
[1]Dept. of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Purpose: To develop and evaluate an automated approach to intelligently interpret free-text, or natural language (NL), physician treatment directives and generate patient-specific prescriptions and clinical goals within a commercial treatment planning system (TPS) using large language models (LLMs), and compare the accuracy, time, and cost between major frontier LLMs, including ChatGPT and Claude.

Methods: We developed a fully-automated multi-agent LLM framework that accepted NL directives as input, extracted prescription and clinical goals through multiple steps of interpretation, and automatically entered them into our TPS (RayStation 2023B).

Results: The mean±standard deviation (SD) percent of correct goals for clinical plans was 82.2±14.4%, attributed to a combination of human error and intentional modification of goals during clinical planning. The mean±SD (p-value) percent of correct goals for 4o-mini, 4o, and Claude was 47.6±26.5% (0.039), 87.5±8.9% (0.236), and 96.0±4.4 (0.043), respectively. Mean±SD inference time per patient was 23.1±12.1 s, 24.2±12.1 s, and 224.7±148.5 s, and mean±SD inference cost per patient was \$0.0042±0.0021, \$0.059±0.031, and \$0.658±0.415, respectively.

Conclusions: An LLM approach for interpreting NL physician treatment directives was developed and demonstrated better agreement with initial directives than final clinical plans. Among LLMs, the most costly reasoning model achieved the highest accuracy and could enable intuitive and efficient auto-planning for complex and unique patients.

CONTACT

Sizhuo Meng
Department of Radiation Oncology, Johns Hopkins University
smeng9@jh.edu
+1 4438971216

INTRODUCTION

Physician treatment directives are commonly documented as free-text instructions, requiring planners to manually identify prescriptions, target coverage requirements, and organ-at-risk (OAR) constraints before entering them into the treatment planning system (TPS). This translation step can be inefficient and vulnerable to interpretation or transcription errors.

Recent advances in large language models (LLMs) provide an opportunity to convert unstructured clinical directives into structured, machine-readable planning inputs.[1,2]

PURPOSE

Free-text physician directives often include heterogeneous prescription details, target coverage requirements, OAR constraints, and patient-specific planning priorities. Manual conversion of these directives into TPS prescriptions and clinical goals requires planner interpretation and may introduce variability.

This study evaluates whether a structured multi-agent LLM framework can convert natural language physician directives into patient-specific, TPS-ready prescriptions and clinical goals for automated treatment planning workflows.

MATERIALS

This study used archived free-text physician treatment directives from five previously treated intensity-modulated proton therapy (IMPT) patients, including GU and CNS cases. Each directive contained patient-specific prescription information and clinical planning instructions written in natural language.

The clinical treatment planning system was RayStation 2023B. The evaluated frontier LLMs included GPT-4o-mini, GPT-4o, and Claude Opus 4. For reference standard construction, an experienced planner manually regenerated clinical goal lists from the archived physician directives, which served as directive-derived ground truth.

METHODS

A fully automated multi-agent LLM framework accepted each natural-language directive as input, interpreted prescription and clinical-goal information through sequential, agent-based processing, and automatically entered the structured outputs into RayStation. LLM-generated goal lists and original clinical plan goal lists were compared with the planner-derived ground truth. Model performance was evaluated by percent correct goals, inference time, and inference cost. Accuracy was compared with clinical plans using paired t-tests.

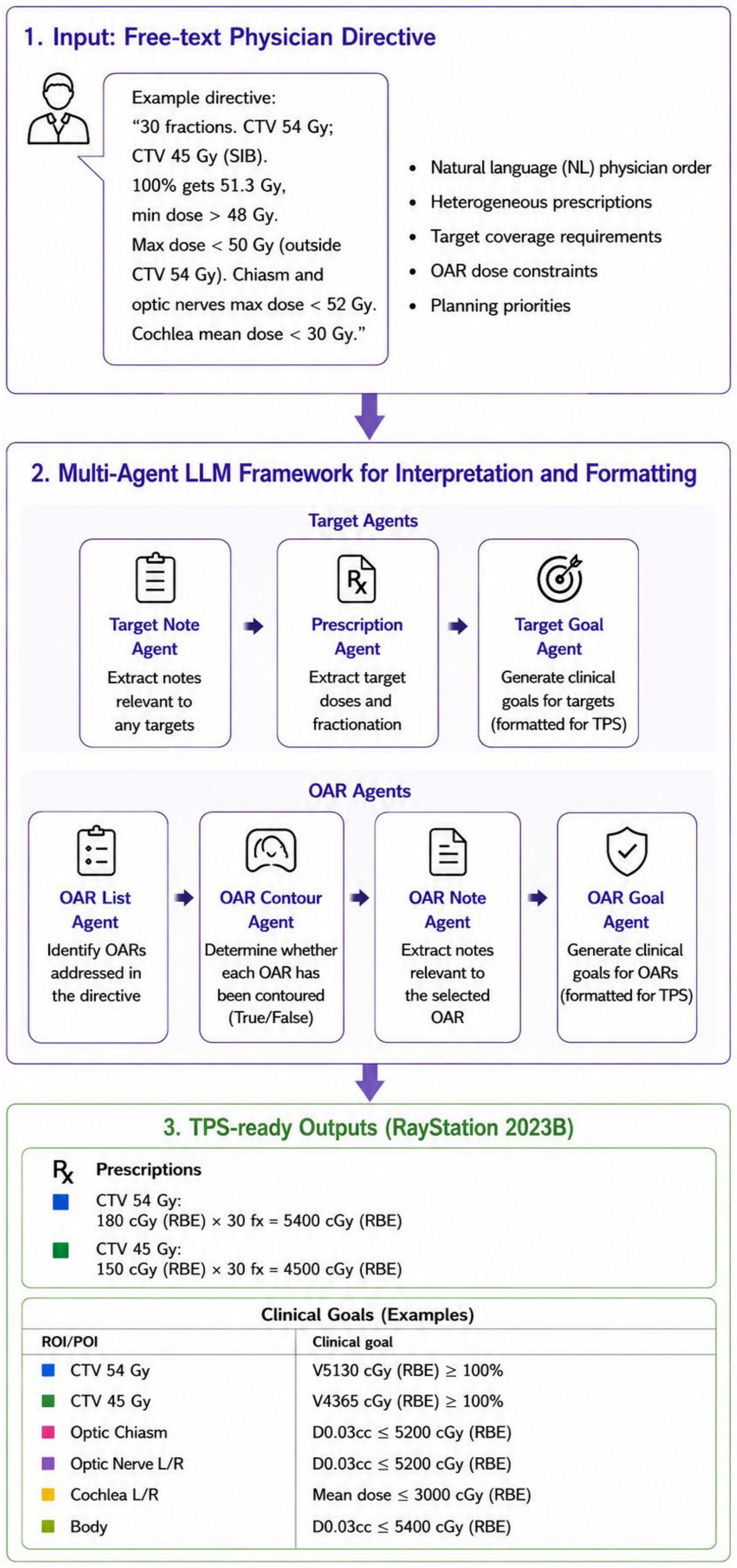

Figure 1. Multi-agent LLM workflow for converting free-text physician directives into TPS-ready prescriptions and clinical goals in RayStation 2023B.

RESULTS

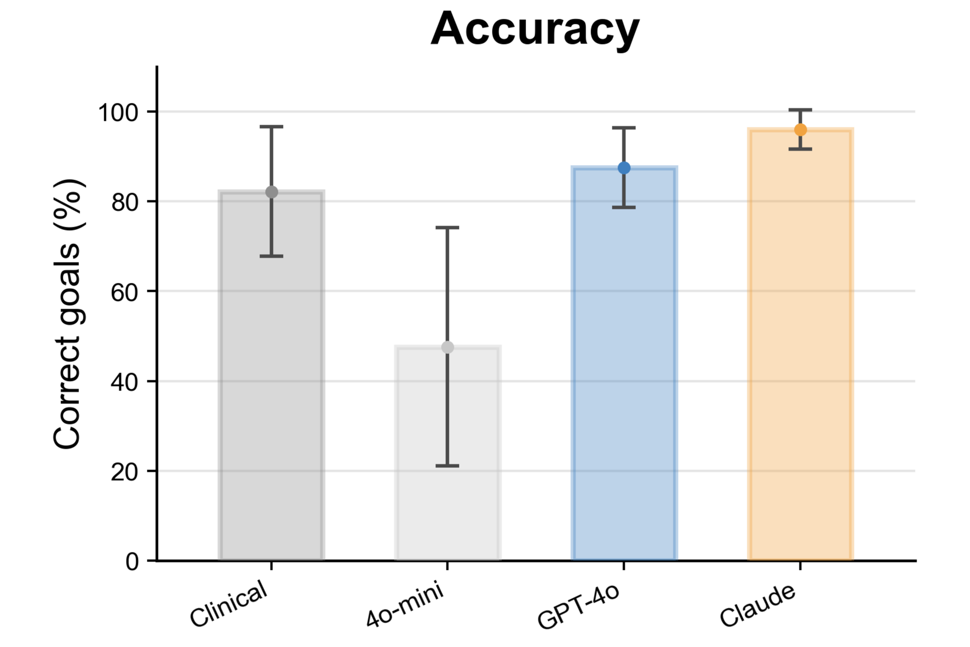

Figure 2. Accuracy of clinical goal agreement with directive-derived ground truth. The percentage of correct clinical goals was calculated by comparing each goal list with the clinical goal list manually regenerated by an experienced planner from the original free-text physician directive. Bars show mean accuracy across five previously treated GU and CNS IMPT cases, and error bars represent ±1 SD.

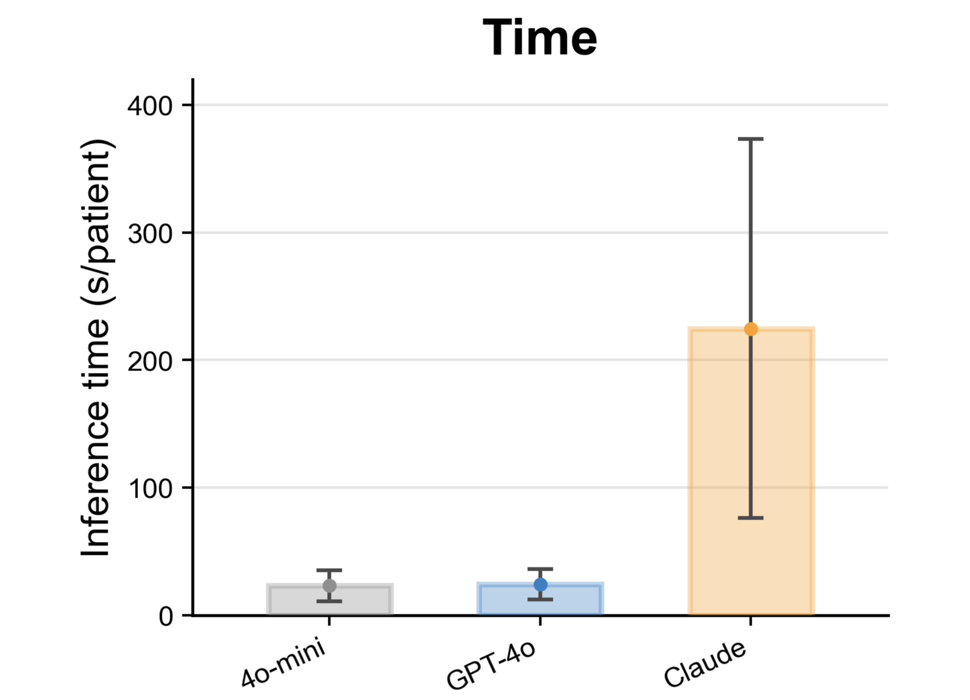

Figure 3. Inference time for automated clinical goal generation across frontier LLMs. Inference time was calculated as the time required for each LLM to interpret the free-text physician directive and generate TPS-ready prescription and clinical goal outputs for each patient. Bars show mean inference time across five previously treated GU and CNS IMPT cases, and error bars represent ±1 SD.

Model	GPT-4o-mini	GPT-4o	Claude
Cost (\$)	0.004 ± 0.002	0.059 ± 0.031	0.658 ± 0.415

Table 1. Inference cost of LLM-based clinical goal generation. Mean inference cost per patient is reported as mean ± SD across five GU and CNS IMPT cases. Cost increased from GPT-4o-mini to GPT-4o and was highest for Claude Opus 4.

CONCLUSIONS

A fully automated multi-agent LLM framework successfully converted free-text physician directives into TPS-ready prescriptions and clinical goals in RayStation. The results suggest that frontier LLMs may reduce manual transcription burden and improve consistency in treatment planning workflows.

REFERENCES

1. Q Wang et al 2025 Phys. Med. Biol. 70 075007
2. Sheng Liu et al 2025 Phys. Med. Biol. 70 155002

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