



Application of a Multi-agent Open-source Large Language Model for Data Abstraction from Radiology Report

Sanaz Vahdati, MD, Mayo Clinic; Pouria Rouzrokh, MD, MPH, MHPE; Elham Mahmoudi, MD, MPH;
Bradley Erickson, MD, PhD, CIIP, FSIIM

Introduction/Background

Recent advancements in large language models (LLMs) have opened new frontiers in artificial intelligence. Multi-agent systems have shown remarkable effectiveness in specialized tasks like data extraction. By emulating collaborative cognitive processes, these systems transform problem-solving approaches, enabling more sophisticated and holistic data analysis. Their ability to distribute complex, multifaceted tasks among specialized agents leads to enhanced decision-making capabilities and more comprehensive solutions. In this regard, we aimed to build a multi-agent language model for radiology report data extraction to investigate their capabilities and assess their performance in deriving specific conclusions from complex medical data.

Methods/Intervention

In this work, we collected 212 radiology reports from two different pathologies. We aimed to extract the presence or absence of acute cervical spine fracture and liver metastasis from radiology reports of the cervical spine and abdominopelvic CT scan collected between January and February 2022. We employed the open-source LLama3-70BInstruct for inference and applied few-shot prompting for each agent. We propose a three-tier multi-agent architecture. This system comprises two verification agents and a reconciliation expert agent, operating in a sequential manner (Figure 1). The initial two agents independently extract data, which is subsequently fed, along with the original report, to the reconciliation agent. This final agent synthesizes the information to produce a comprehensive conclusion. We evaluated the efficacy of this pipeline through performance metrics.

Results/Outcome

The multi-agent model for liver metastases assessment demonstrated high performance, achieving accuracy of 0.95, F1 score of 0.95, Positive Predictive Value (PPV) of 0.96, and Negative Predictive Value (NPV) of 0.95. Our model could exclude patients with a prior history of metastasis from new diagnosis classification. For the extraction of acute cervical spine fracture presence, the pipeline exhibited robust performance metrics: accuracy of 0.96, F1 score of 0.92, PPV of 0.92, and NPV of 0.97. In both tasks, the reconciliation agent provided salient cues pertaining to the final determination, facilitating further elucidation of results.

Conclusion

In conclusion, we built a multi-agent orchestration using debating scenarios to boost collective reasoning of data abstraction from the radiology report.

Statement of Impact

Multi-agent models with the potential to collective intelligence bear considerable promise for data extraction from radiology reports.

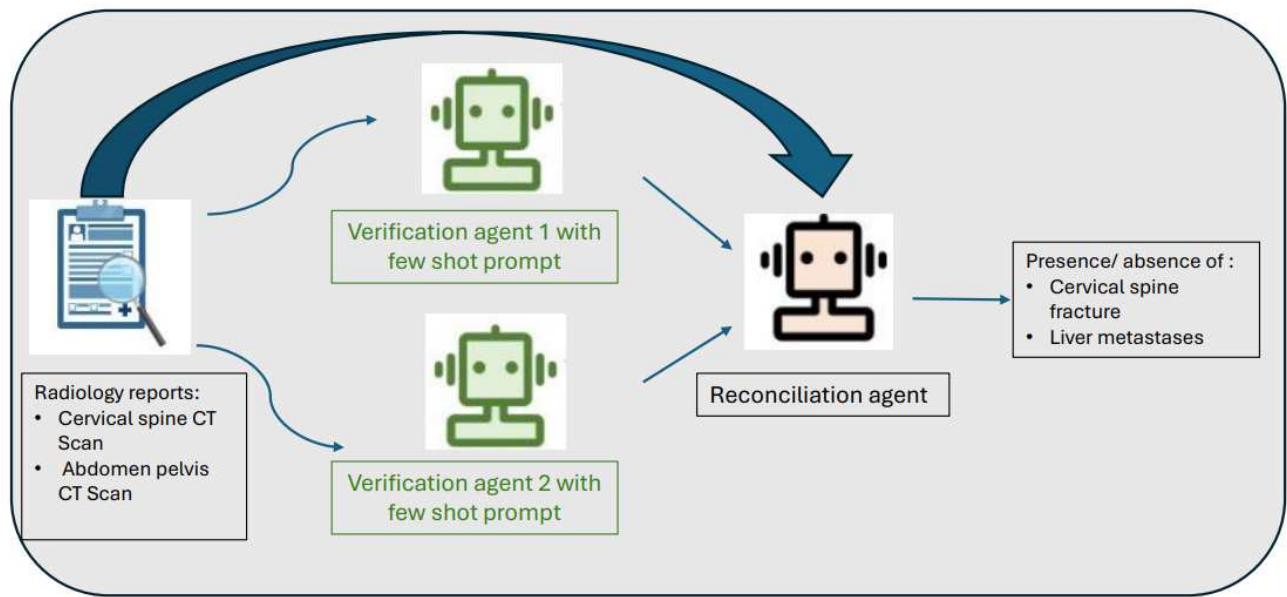


Fig1. Schematic of Multi-Agent architecture for radiology report data extraction.

Keywords

Radiology report; Large language models; Multi-agent