



# Developing a Standardized Ontology for Imaging Findings Using Large Language Models with Crowd-Sourced Iteration and Validation

Yilun Zhang, MD, Transitional Year Resident, RWJBH Community Medical Center/Massachusetts General Hospital Tarik Alkasab, MD, PhD

## Background/Problem Being Solved

Radiology reports contain valuable imaging findings and clinical recommendations, but inconsistent reporting styles and lack of a standardized structure and ontology hinder their integration into clinical workflows. A reliable, comprehensive, and consistent 'lingua franca' for imaging findings with extractable structure and features allows for bidirectional automation of report generation and understanding.

#### Intervention(s)

In collaboration with the Radiological Society of North America (RSNA) and the American College of Radiology (ACR), we developed and leveraged the Common Data Elements (CDE) framework to evaluate and expand standardized radiology ontologies. A large language model (LLM) pipeline was developed using de-identified local report texts from Massachusetts General Hospital and publicly available datasets such as MIMIC, enabling the creation and refinement of a robust, scalable framework.

## Barriers/Challenges

Key challenges included harmonizing diverse reporting styles and achieving consensus for CDE definitions among radiologists in a scalable, federated manner without exposing patient information.

#### Outcome

We developed an open-source, LLM-powered tool that enables users to access validated CDEs, contribute new or iteratively revise existing CDEs with the aid of an LLM, and ingest their local reports for analysis without risk of data exposure. The tool also unifies across and integrates with RadLex, SNOMED CT, LOINC, UMLS, and other ontologies. This allows for distributed and secure updates to the CDE framework and LLM-powered utilities in a scalable and asynchronous manner.

## Conclusion/Statement of Impact/Lessons Learned

This work demonstrates the potential of an open-source tool that combines LLMs and the CDE framework to create a standardized, scalable ontology for radiology findings with additional utilities such as LLM-generated CDEs and reports and automatic extraction of CDEs from a local corpus of reports. This novel framework and tool drives human-in-the-loop

automation, fosters interoperability, and revolutionizes clinical decision-making by bridging the gap between fragmented data silos and transforming the traditionally free-text narrative report into a structured output of actionable clinical insights.

#### Keywords

Applications; Artificial Intelligence/Machine Learning; Clinical Workflow & Productivity; Emerging Technologies; Quality Improvement & Quality Assurance; Standards & Interoperability