SIIM CMIMI23
OCTOBER 1-2
Johns Hopkins | Baltimore, MD
Conference on Machine Intelligence in Medical Imaging

Using An Open-source Language Model to Abstract the Presence of Acute Cervical Spine Fracture from Radiologic Reports: A HIPAA Compliant Alternative to "ChatGPT"

Bardia Khosravi, MD, MPH, MHPE, Postdoctoral Research Fellow, Mayo Clinic AI Lab; Sanaz Vahdati, MD; Pouria Rouzrokh, MD, MPH, MHPE; Shahriar Faghani, MD; Mana Moassefi, MD; Ali Ganjizadeh, MD; Bradley J. Erickson, MD, PhD, CIIP, FSIIM

Introduction

ChatGPT is a large language model (LLM) with a chat interface that allows for back-and-forth communication with users. While the engineering behind aligning the model with human values is a significant advantage, the closed-source nature of the model and the need to transmit sensitive patient information through a commercial API make it unsuitable for medical data abstraction. Additionally, all studies that have used ChatGPT to extract radiologic reports have created synthetic reports to comply with HIPAA regulations. As a result, open-source LLMs may be a more viable option for data abstraction from radiologic reports because they can be deployed locally.

Hypothesis

In this study, we use an open-source LLM to determine the presence and level of cervical spine fractures in radiologic reports.

Methods

We collected radiologic reports from 115 non-contrast cervical spine CT scans taken between January and February 2022. Of these reports, 15 were used for prompt engineering and fine-tuning, while the remaining 100 were used to test the model's performance. The model's objective was to generate a JSON file indicating the presence or absence of an acute cervical vertebral fracture and its associated level. This particular problem was chosen because the model must distinguish between different anatomical regions and differentiate between chronic and acute fractures. We used an open-source LLM, specifically the Large Language Model Meta AI (LLaMA-13B variation), and deployed it locally. We employed a combination of chain-of-thought (CoT) and few-shot prompting techniques, asking the model to produce a final JSON output of the results (Figure 1).

Results

The model consistently produced correctly formatted JSON outputs. It achieved a 99% (99/100) accuracy rate in detecting acute cervical spine fractures. Notably, in 16 out of 17 cases where fractures were present, the model accurately predicted the level of the fracture. However, in five instances where multiple vertebrae levels were fractured, the model only predicted one.

Conclusion

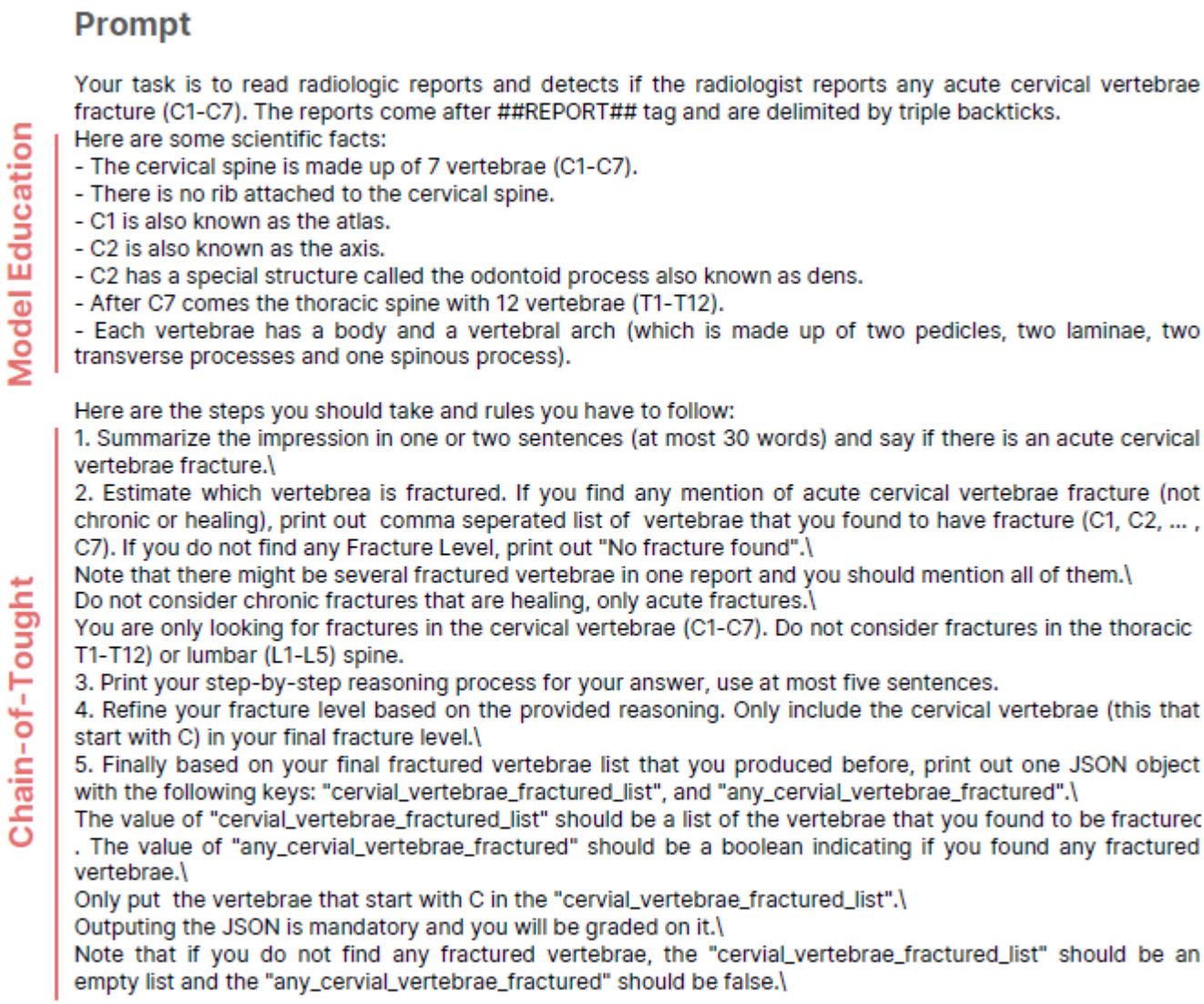
With appropriate prompting, open-source LLMs can perform well without the need to share patient data with

external sources.

Statement of Impact

LLMs can be used to extract information from unstructured text. While ChatGPT has gained significant popularity, it only offers a commercially hosted API, which requires transmitting patient information outside of the healthcare institution. It is important to explore open-source alternatives and their capabilities in sensitive scenarios involving protected health information (PHI).

Figure 1. The final prompt to the model.



Output format constraint

here is how you should format your output:\
<Summary>
* your summary *
</Summary>
<Initial Fracture Level>
* list of Fracture Level or if none type No fracture found *
</Initial Fracture Level>
<Reasoning>
* your reasoning *
</Reasoning>
<Final Fracture Level>
* your final Fracture Level *
</Final Fracture Level>
<JSON>
* your JSON output *
</JSON>

Few-shot example 1

~~~
## REPORT ##
```1. Acute type II odontoid fracture with posterior displacement of the odontoid and associated prevertebral hematoma. No significant spinal canal narrowing. 2. Multilevel, moderate spondylotic changes of the cervical spine. ```

<Summary>
There is an acute type II odontoid fracture.
</Summary>
<Initial Fracture Level>
C2
</Initial Fracture Level>
<Reasoning>
Based on the report, there is an acute type II odontoid fracture in the report. The odontoid is part of C2. So C2 has an acute fracture.
</Reasoning>
<Final Fracture Level>
C2
</Final Fracture Level>
<JSON>
{
  "any_cervial_vertebrae_fractured": true,
  "cervial_vertebrae_fractured_list": ["C2"]
}
</JSON>
~~~

Few-shot example 2

~~~
## REPORT ##
```1. Abnormal widening of the right C2-C3 facet joint without definite associated fracture. Findings are concerning for possible underlying associated ligamentous injury. 2. Advanced spondylotic changes throughout the cervical spine. ```

<Summary>
There is no mention of acute cervical vertebrea fracture.
</Summary>
<Initial Fracture Level>
No fracture found
</Initial Fracture Level>
<Reasoning>
Based on the report, there is a mention of abnormal widening of the right C2-C3 facet joint without definite associated fracture, but there is no acutely fractured cervical vertebrae.
</Reasoning>
<Final Fracture Level>
No fracture found
</Final Fracture Level>
<JSON>
{
  "any_cervial_vertebrae_fractured": false,
  "cervial_vertebrae_fractured_list": []
}
</JSON>

Question

~~~
REPORT
```%%REPORT%%```

<Summary>

Keywords
Natural Language Processing; Large Language Models; Privacy; Open Science
```