



## 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.

Model Education

Chain-of-Thought

### Prompt

Your task is to read radiologic reports and detects if the radiologist reports any acute cervical vertebrae fracture (C1-C7). The reports come after `##REPORT##` tag and are delimited by triple backticks.

Here are some scientific facts:

- The cervical spine is made up of 7 vertebrae (C1-C7).
- There is no rib attached to the cervical spine.
- C1 is also known as the atlas.
- C2 is also known as the axis.
- C2 has a special structure called the odontoid process also known as dens.
- After C7 comes the thoracic spine with 12 vertebrae (T1-T12).
- Each vertebrae has a body and a vertebral arch (which is made up of two pedicles, two laminae, two transverse processes and one spinous process).

Here are the steps you should take and rules you have to follow:

1. Summarize the impression in one or two sentences (at most 30 words) and say if there is an acute cervical vertebrae fracture.\
2. Estimate which vertebrae is fractured. If you find any mention of acute cervical vertebrae fracture (not chronic or healing), print out comma separated list of vertebrae that you found to have fracture (C1, C2, ... , C7). If you do not find any Fracture Level, print out "No fracture found".\

Note that there might be several fractured vertebrae in one report and you should mention all of them.\

Do not consider chronic fractures that are healing, only acute fractures.\

You are only looking for fractures in the cervical vertebrae (C1-C7). Do not consider fractures in the thoracic T1-T12) or lumbar (L1-L5) spine.

3. Print your step-by-step reasoning process for your answer, use at most five sentences.
4. Refine your fracture level based on the provided reasoning. Only include the cervical vertebrae (this that start with C) in your final fracture level.\
5. Finally based on your final fractured vertebrae list that you produced before, print out one JSON object with the following keys: "cervial\_vertebrae\_fractured\_list", and "any\_cervial\_vertebrae\_fractured".\

The value of "cervial\_vertebrae\_fractured\_list" should be a list of the vertebrae that you found to be fractured . The value of "any\_cervial\_vertebrae\_fractured" should be a boolean indicating if you found any fractured vertebrae.\

Only put the vertebrae that start with C in the "cervial\_vertebrae\_fractured\_list".\

Outputting the JSON is mandatory and you will be graded on it.\

Note that if you do not find any fractured vertebrae, the "cervial\_vertebrae\_fractured\_list" should be an empty list and the "any\_cervial\_vertebrae\_fractured" should be false.\

## Output format constraint

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

## Few-shot example 1

~~~\n## REPORT ##\n```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. ```\n\n<Summary>\nThere is an acute type II odontoid fracture.\n</Summary>\n<Initial Fracture Level>\nC2\n</Initial Fracture Level>\n<Reasoning>\nBased 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.\n</Reasoning>\n<Final Fracture Level>\nC2\n</Final Fracture Level>\n<JSON>\n{\n "any\_cervical\_vertebrae\_fractured": true,\n "cervical\_vertebrae\_fractured\_list": ["C2"]\n}\n</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 vertebra 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_cervical_vertebrae_fractured": false,
  "cervical_vertebrae_fractured_list": []
}
</JSON>
```

## Question

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

<Summary>
```

## Keywords

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