



Large Language Model Integration with Radiology Report Query Tool Enables Complex Searches

Kurt Teichman, M.Eng, Senior Programmer/Analyst, Radiology, Weill Cornell

James Ledoux, PhD; Chenglin Zhu; Keith Hentel, MD, MS; Jacob Kazam, MD; Martin Prince, MD, PhD; Shih George, MD, MS

Background/Problem Being Solved

Most radiology report search tools (eg, document storage databases like solr and elastic search combined with basic NLP) allow for useful queries of radiology reports (eg, Find all the positive cases of appendicitis on CT studies in the ED for the last 12 months). However, they are limited in their capacity to do data extraction and/or filtering on documents for more complex requirements (eg, extracting discrete measurements out of free text in a structured format like JSON and/or complex queries like if lesions measure ≥ 2.5 cm).

Intervention(s)

Here we provide methodology for integrating large language models (LLMs) on top of radiology report an initial query (eg, renal cell carcinoma on CT studies), and allow a user-defined LLM prompt ('Find all renal tumors ≥ 2.5 cm and provide in JSON format') to be run on the initial query result. Our approach uses a template mechanism, composing a report alongside a user prompt, to be sent to the LLM to be processed. This is similar to typical retrieval augmented generation (RAG) templating. We also provide means of storing query history and persisting LLM results as they can often take a long time to finish.

Barriers/Challenges

Accessing a LLM (e.g. OpenAI ChatGPT) is often difficult from inside institutions following HIPAA compliance, and thus, running your own LLM (e.g Llama 3.1 8b) is ideal. Providing multiple users reliable access to LLM services is also challenging as result processing takes time (i.e. each report needs to be passed in with your query). Asynchronous methodology is required. We utilize 2 queues, redis based as well as a queue provided with vLLM.

Outcome

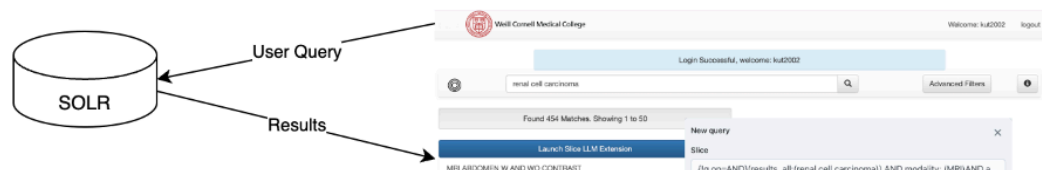
Users are able to obtain refined search results often saving them considerable amounts of time. We should note, however, that these results are "mostly" accurate. The usage of smaller language models (e.g Llama 3.1 8B) can sometimes hallucinate results when input is of longer context size

Conclusion/Statement of Impact/Lessons Learned

We leverage LLMs along with modern software engineering practices to allow users to obtain more granular results searching radiology report archives. Future work will incorporate Multimodal LLMs to provide additional options for queries.

Figure(s)

1.)



2.)

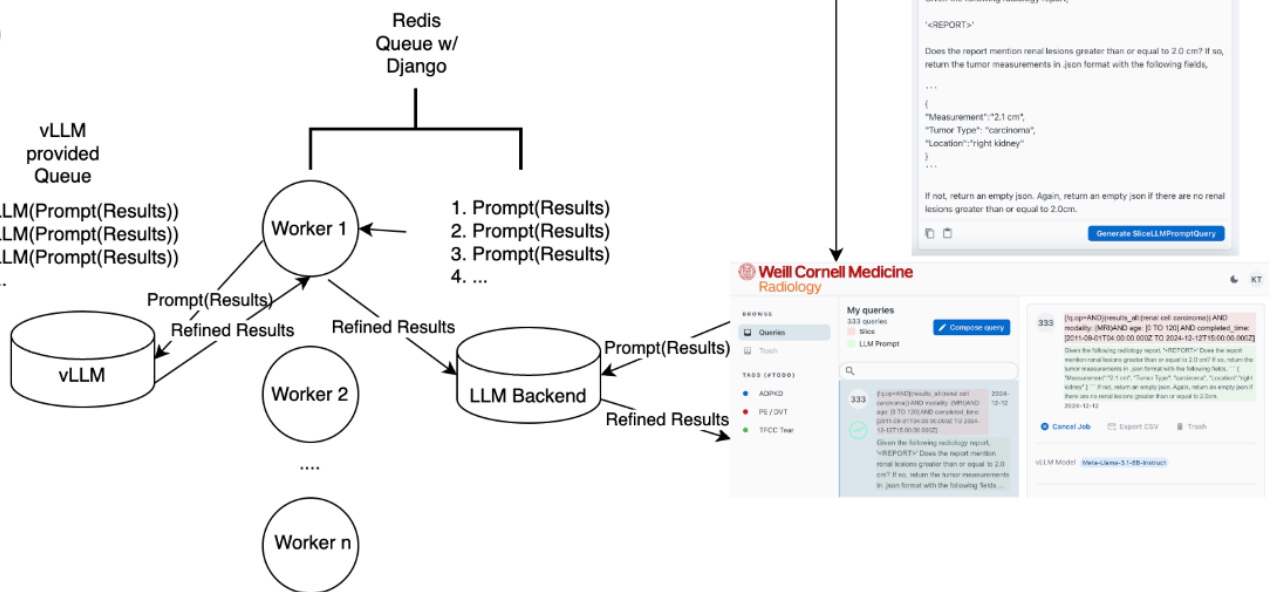


Figure 1 - Step 1.) the user generates a keyword search +/- date search +/- other fields (e.g. radiologist, modality. Step 2.) Results are sent to the Report Query Tool LLM extension which allows users to create a template containing their report results as well as a LLM prompt to be applied to the results.

Example prompt templates

Figure 1. Example Report + Prompt Templates

Example prompt+report templates

Given the following radiology report,

'<REPORT>'

Given this radiology report, '<REPORT>' Please give the height of the patient in meters. Please give the weight of the patient in kg. Please tell me if there is a comparison available? If so, what is the date? Please give me the mayo classification. Please give me the liver fat fraction. Please give me the liver volume. Please give me the HtLV, if it exists. Please give me the left kidney volume. Please give me the previous left kidney volume, if it exists. Please give me the number of cysts in the left kidney, if any. Please give me the cyst fraction of the left kidney, if it exists. Please give me the largest cyst of the left kidney, if it exists. Please give me the right kidney volume. Please give me the previous right kidney volume, if it exists. Please give me the number of cysts in the right kidney, if any. Please give me the cyst fraction of the right kidney, if it exists. Please give me the largest cyst of the right kidney, if it exists. Please give me the Total kidney volume (TKV), if it exists. Please give me the HtTKV, if it exists. Output these answers in a JSON object.

...

Given the following radiology report,

'<REPORT>'

Does the report mention renal lesions greater than or equal to 2.0 cm? If so, return the tumor measurements in .json format with the following fields,

```
{  
  "Measurement": "2.1 cm",  
  "Tumor Type": "carcinoma",  
  "Location": "right kidney"  
}
```

If not, return an empty json. Again, return an empty json if there are no renal lesions greater than or equal to 2.0cm.

...

Figure 2. Software Architecture associated with the Report Query + LLM deployment

Keywords

Artificial Intelligence/Machine Learning