

## Clinical LLM Selection Tool Usage Guidelines

### Step 1: Determine the Clinical Workflow Stage

1. Identify the clinical workflow stage in which you want to use Large Language Models (LLMs) or the stage that matches the tasks you wish to perform with LLM assistance.
  - Stage I: registration and department guidance
  - Stage II: pre-diagnosis and examination
  - Stage III: diagnosis and treatment planning
  - Stage IV: treatment and hospitalization
  - Stage V: discharge and follow-up

Example: If you are a radiologist looking to use LLMs for writing radiology reports, select Stage II. If you are a clinician interested in using LLMs to assist with patient diagnosis, select Stage III.

### Step 2: Complete the Questionnaire Based on Your Needs

1. After selecting the stage, answer the questions in the tool to receive model recommendations for your use case.
2. If you see underlined text in blue in a question, it provides additional explanations when you click on it. The explanation will appear at the bottom of the page and can be closed by clicking the red 'close' button.

#### Example Scenario:

<p>Question 1:</p> <p>Does your input include only <a href="#">textual data</a>?</p> <p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No</p>	<p>Question 4:</p> <p>Do you need LLM to be able to <a href="#">generate text related to radiology</a>?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>
<p>Question 2:</p> <p>Does your output include only <a href="#">textual data</a>?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Question 5:</p> <p>Do you need LLM to be able to <a href="#">summarize text related to radiology reports</a>?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>
<p>Question 3:</p> <p>Do you agree that your data may be shared with third parties, published, or made generally available?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Question 6:</p> <p>Do you need LLM to be able to <a href="#">process medical images</a>?</p> <p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No</p>
	<p>Question 7:</p> <p>Do you need LLM to be able to <a href="#">answer questions related to examination</a>?</p> <p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No</p>

- User Role: Radiologist
- Intended Stage of Use: Stage II (assisting with radiology exams and report generation)
- Sample Answers:
  - Question 1: No (for radiology reports that include medical images)
  - Question 2: Yes (to generate text related to radiology reports)
  - Question 3: No (when data includes patient personal information)
  - Questions 4-7: Choose based on your specific requirements; for instance, “Yes” for question 4, “Yes” for question 5, “No” for question 6 and “No” for question 7.

## Deployment Choice (Question 8)

Question 8:

Do you prefer to deploy LLM [locally](#) rather than [in the cloud](#)?

☒ Yes ☐ No

The LLMs we recommend you use are: RaDialog-7B (\*), PaLM-E-84B, LLaMA2. Here are the required resources:

RaDialog-7B: fine-tuning requires 1 NVIDIA A40-48GB GPU. The total thermal design power of the GPU is 300W and the total price is \$7799.

It can be accessed via link [RaDialog GitHub](#).

PaLM-E-84B: no specific resource requirements for this model, refer to resources for LLMs of similar size (70B). Fine-tuning requires 8 H100 GPUs. The total thermal design power of the GPU is 2400W and the total price is \$36000.

LLaMA2: the exact model size is not mentioned in the papers and is assumed here to be LLaMA2-7B. Fine-tuning requires 1 NVIDIA Quadro RTX 8000-48GB GPU. The total thermal design power of the GPU is 295W and the total price is \$4300.

Question 8:

Do you prefer to deploy LLM [locally](#) rather than [in the cloud](#)?

☐ Yes ☒ No

The LLMs we recommend you use are: RaDialog-7B (\*), PaLM-E-84B, LLaMA2. Here are the required resources:

RaDialog-7B: fine-tuning requires 1 NVIDIA A40-48GB GPU. The price of renting a GPU with 48GB of memory using a cloud service is \$3.75/1h.

It can be accessed via link [RaDialog GitHub](#).

PaLM-E-84B: no specific resource requirements for this model, refer to resources for LLMs of similar size (70B). Pre-train requires more than 120 NVIDIA TESLA A100-80GB GPUs. The price of renting 120 this type of GPU using cloud services is \$750/1h.

Fine-tuning requires 8 H100 GPUs. The price of renting 8 this type of GPU using cloud services is \$88/1h.

LLaMA2: the exact model size is not mentioned in the papers and is assumed here to be LLaMA2-7B. Fine-tuning requires 1 NVIDIA Quadro RTX 8000-48GB GPU. The price of renting a GPU with 48GB of memory using a cloud service is \$3.75/1h.

This question offers options for local deployment and cloud deployment, depending on available resources. Generally, we recommend cloud services for pre-training or fine-tuning, as they can reduce computing and cost demands. For the inference phase, local deployment is typically more cost-effective.

## Step 3: Review the Recommended LLMs

1. After completing the questionnaire, the tool will display, if available, a list of LLMs that meet your requirements, along with specific details such as:

- Model Name (+ Size)
- GPU Memory Requirements
- Energy Consumption
- Pricing Information
- Access Links (e.g., GitHub)

2. Revising Answers: You can go back to any question and adjust your answers. Changes will update the following questions and recommendations.

3. Performance Indicator (\*): If an LLM name is followed by an ‘(\*)’, it indicates that multiple papers have reported this LLM as best performing for certain tasks. This star represents high performance based on research findings, though this may apply across various clinical contexts.

## Disclaimer

*The data provided within this selection tool is derived from reported research findings in collected papers and is intended solely for reference in the selection of LLMs. The authors and developers of this tool do not assume any responsibility or liability for the outcomes arising from the use of this tool, including clinical, financial, or legal implications. This tool does not constitute medical or professional advice. Users are encouraged to verify all information independently and consult relevant medical or technical experts before implementing any LLM recommendations. Prices may vary from different regions and may change in the future. Please refer to your provider for potential price updates. The authors are not liable for discrepancies.*