
RAG vs No RAG Large Language Models

The Mission: To understand the strengths of RAG systems and interact with them.

The Resources: You will be provided with 2 links but you need to find a CV or generate one with an LLM and humanify it.

The link to task 1-2:

<https://huggingface.co/spaces/AnasKAN/Bart-RAG-vs-NoRAG-comparison>

Task 1: CV summarization (RAG with specialized LLM)

Goal: Analyze and note the differences between RAG vs no RAG.

You will use task specific LLM called Bart, which is tuned for summarization tasks. You will specialize it into the domain of your expertise if possible (you can do that by providing a CV from the internet or generate a CV from an LLM and then humanify it).

Note your findings:

1. **Diagnosis:** What are the differences of **No RAG** vs **RAG** in terms of response length, how accurate it is, and other aspects
2. **Conclusion:** After experimenting with this LLM, what do you **conclude**?

Note:

Task 2: Instruction following

Goal: Try your best at making BART follow your instruction.

In task 1, you have seen the difference in performance between RAG and without. Now we want you to make it summarize in a style you like (e.g. summarize in 10 words)

Note your findings:

1. **The analysis:** Is it following your instructions, or not?
2. **Reasoning:** Note why, what do you think the reasons are?

Notes:

The link to task 3-4:

<https://huggingface.co/spaces/AnasKAN/Qwen2.5-0.5B-RAG-vs-NoRAG>

Task 3: RAG with general LLM

Goal: Specialize the general LLM with RAG and without in default settings, and note your conclusion.

You will interact with a general LLM called Qwen2.5, a 0.5B model. Use the default settings when doing this challenge.

Write a file to specialize the LLM (you can use an LLM to generate that file and then humanify it), then ask it specialized questions.

Note your findings:

3. **The analysis:** What are strengths and weaknesses of RAGs and without?
Try to find 3.
4. **The conclusion:** conclude what are the cases that will make you use RAG and when will you use No RAG?

Notes:

Task 4: Controlling the generation

Goal: Experiment with generation settings.

As you might have already seen, there are those sliders, namely: Temperature, Top-P, Repetition Penalty, Max New Tokens. Experiment with those, understand what each setting does, then try to find the best setting.

The Questions are:

5. **The analysis:** What each setting does?
6. **Importance:** Determine what setting affects the output the most?

Notes:

Task 5: Advisor agent

Goal: make your own tailored AI Agent in PIF AI space.

Enter the PIF AI space site, and from what you learned from past tasks, apply your knowledge to create an AI Agent.

The Questions are:

7. **The analysis:** Note the quality of answer from your specialized model vs a general model in the AI space.
8. **Share:** tell us what kinds of uses this will benefit you in your work? (e.g. you get a second opinion on a matter before asking the human expert in the problem you face)

Notes:
