LLM Assignment-1 :: *Debnath Kundu (MT22026)*

- Show 5 examples of hallucination in Self-consistency, Fact-checking, and summarization each using the above LLMs. 15 examples per LMM (30 in total). Use RAG to solve/minimize hallucinations.

Hallucinations in LLMs mean that they generate information not present in their training data or the input provided to them. These hallucinations can range from fabricating information to inaccuracies, posing challenges to their reliability in critical applications. Retrieval-augmented generation (RAG) significantly reduces hallucinations in LLMs by combining the generative capabilities of the model with information retrieval (IR) systems.

In the ipynb file for **LLAMA2**, I have compared 5 cases of Fact-checking, Self-consistency and summarisation hallucinations and the output with RAG.

Fact-Checking inference:

- The model struggles to retrieve facts that have happened after its release. (`Chandrayaan-3`)
- Moreover, RAG also improves the accuracy/conciseness of the outputs. (`Backpropagation`, `Covid-19`)

Self-Consistency inference:

- The model struggles to answer if the same query is slightly twisted. (`Newton's law`)
- In descriptive queries, the model lost accuracy, i.e., it brings about any knowledge it has about the query terms. It seems more of a BagOfWords model! RAG significantly improves in this scenario. (`Radioactivity`)
- The model also needs help to recognise if some past information is updated. (`TATA IPL`)
- If some higher-level logical question is asked, which is not present in that particular form in the training corpus, then the model hallucinates. (`Can energy be destroyed?`)

Summarisation inference:

- The model fails to exemplify any scientific concept. (`concept of infinity`)
- It also needs to answer descriptive logical questions if they have any addendum. (`How do vaccines work to prevent diseases?`)
- If a term match has some definition in the training data, then the model can summarise it well. RAG only improves a little here. (`Deep Convolutional Networks`)
- If the query term doesn't have any definition, then the model returns the matching text snippets. RAG helps to identify the relevant sources. (`RNN Encoder-Decoder`)