

Exercise: Evaluation of Retrievers

- **Background:** The company has asked you to create a system where a user gives a context and a question and the system should generate the answer on the basis of the context
- **Type of Data:** Unstructured textual data
- **Data:**
 - You need to use the squad dataset
Download the dataset and read the readme file at:
squad_v2_sampled dataset in the Datasets section of classroom
- **Expected Result:** The retrieval chunks should be
- **Settings/Changes that need to be experimented:**
 - Retrieve top_k (with different top_k values like 1, 3, 5, 10, 20) contexts and calculate the retrieval metrics. Then pass these context to LLMs/model to extract the results and then calculate answer metrics
 - Use different algorithms like Keywords (TF-IDF , BM25 etc) , Vector Search (Use different embedding models) , and Hybrid Search (Keyword + Vector).
 - See if any of the optimization techniques post retrieval improve the results.
 - You can use following rerankers:
<https://huggingface.co/ibm/re2g-reranker-trex>
<https://www.sbert.net/docs/pretrained-models/ce-msmarco.html>
 - Add the retriever and reranker details inside the model details section of your report
 - You can use any open source library to store the contexts and retrieve them
- **Libraries to use:** langchain, FAISS chromaDB, etc

Deliverables

- Do the assignment of google colab or your laptop.
- Notebook containing the codes should be uploaded to github
- Provide a report (report can be txt or docx files) with following details:

- Generate retrieval metrics like recall@K , MRR, NDCG etc. for each combination of algorithm and top k value.If any optimization technique is used , add a comparison of before and after.
Note: include all the metrics necessary for the task
- Upload the colab notebook and the report on github inside the subfolder named "Assignment 2"