**Adding this as reference**

Project Proposal : Question generation is less popular when compared to Question Answering task amongst the NLP applications. The project goal is to build an end to end Question Answer generation system that could take a context as input and generate meaningful question answer pairs related to the context using the huggingface pre-trained transformer models like FLAN-T5, keyBERT etc. This is a 2 step process - Identify keywords from the context and then use a seq2seq model (T5/flan-T5) to generate meaningful questions. As part of this build, analyze the previous work done in this area and try to come up with **some alternatives/improvements.** Educators spend a lot of time manually preparing the content for assessments and the proposed automation could be a huge benefit by reducing the manual effort spent on these tasks.

1. Goal

Given the context , generate Question Answer pairs

1. Question and Answer generation

Explain QA Task Variants

There are different QA variants based on the inputs and outputs:

Extractive QA: The model extracts the answer from a context. The context here could be a provided text, a table or even HTML! This is usually solved with BERT-like models.

Abstractive QA: <>

**Based on Research Paper: https://arxiv.org/pdf/2005.01107v1.pdf**

Other implementations and ideas:

<https://github.com/patil-suraj/question_generation> (End to End QA generation model based on the above research implementation )

<https://www.youtube.com/watch?v=hoCi_bJHyb8>

1. What’s new or different with this implementation?

* End to end Question answer generation is not a well explored area
* Used latest hugging face transformer models like flan-t5 to come up with the E2E Question answer generation system based on the logic outlined in the research paper
* Generates Multiple Question Answer pairs for the given context
* Tried to apply the logic for a dataset related to Edtech domain

1. Models used in this implementation

KeyBert , Yake, **T5 (Explain the architecture)** , **Flan-T5** (Latest: <https://huggingface.co/google/flan-t5-xxl>)

, Haystack QA framework (used this to preprocess biology textbook data)

1. Implementation

Flow diagram explaining the implementation logic

1. Evaluation Metrics <tbd>
2. Future improvements

Fine tune flan-t5 for E2E Question answer generation

Current:

Generate Question: “Context”

Future:

Generate question and answer: “Context”

And generate possible QA pairs (Both abstractive and extractive)

Instead of generating all possible QA pairs in a literal sense, generate an exhaustive amount that appears like its all possible QA pairs