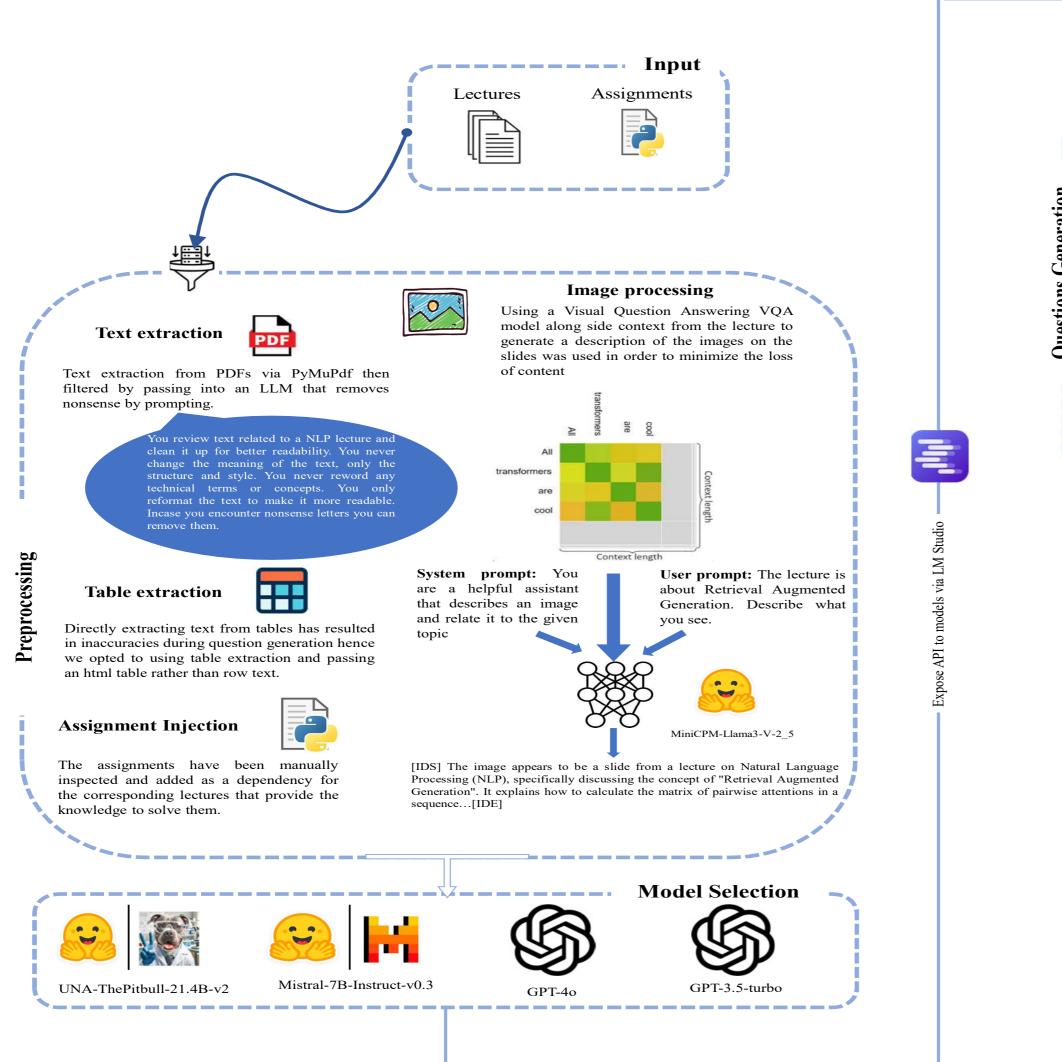
Automatic Exam Question Generation (Lecture2Exam)

Mohammad Saknini, Darius Arbabha, Shinas Shaji

Motivation

Creating exam questions is tedious for the professors and often time-consuming, detracting from their primary responsibilities of teaching and research. Developing a large language model (LLM) pipeline to generate exam questions from lecture notes aims to streamline this process, enhancing efficiency and accuracy. This automated solution ensures comprehensive coverage of course material, producing high-quality, consistent questions. It allows customization to suit different question formats and difficulty levels. Ultimately, this pipeline should enable educators to focus more on student engagement and the quality of instruction, rather than stressing about the exam questions.

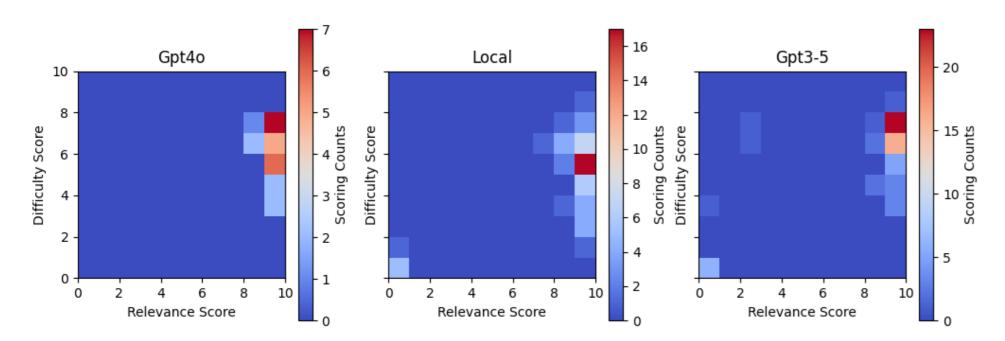
Workflow



For a consistent more reliable output the questions are generated iteratively which reduces the complexity of the request but require more tokens due to having to pass the context to the model times for there different calls using diverse in nature across the slides. Restrict the questions to **Prompt** How many free text questions? How many coding Free Text Prompt Code Question MC Questions Prompt Prompt MC Questions Free Text Questions **Questions** Pool **Coding Questions** The generated questions are evaluated for quality in terms of metrics like relevance to the lecture and the difficulty of the question in context of the lecture. The overall questions coverage of the lecture is also evaluated. The evaluation model is also prompted to reason for these metrics, and the answer to the question is also generated from the lecture context. given the lecture content within <lecture> </lecture> and question within <question> </question> tags. Always provide System a response in the following format using the appropriate Prompt reasoning for the evaluation in markdown text and close with</reasoning> <relevance>an integer from 0 to 10.. Question Prompt Lecture Context Difficulty Score Overall LLM Reasoning **Evaluation** Relevance Score Coverage LLM Answer

Results

Questions distribution by Relevance and Difficulty for each model



Contact

Saknini Mohammad, Darius Arbabha, Shinas Shaji Hochschule Bonn-Rhein-Sieg Email: firstname.lastname@smail.inf.h-brs.de



Acknowledgement

Github Repository

Thanks to Prof. Dr. Jörn Hees and Tim Metzler for the materials and the insightful lecture **Natural Language Processing**.

We want to also thank Fraunhofer INT for allowing us to use an OpenAl API key for this project.





Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Survey

Group 1

- What is the importance of adding special tokens like [CLS], [PAD], [SEP], and [MASK] during the
 preprocessing of text data for transformer models.
- What is a Retrieval Augmented Generator (RAG) system, and how does it help in reducing the context size for document retrieval and generation tasks?
- What is the purpose of using Term Frequency Inverse Document Frequency (TFIDF)?

Group 2

- Explain how one-hot encoding works and provide an example using a simple sentence.
- How do transformers solve the problem of parallelization in sequence-to-sequence models and why is this significant for NLP tasks?
- What is the primary purpose of training language models with human feedback?

Group 3

- How does prefix tuning differ from parameter-efficient fine-tuning methods like LoRa, in terms of their approach to updating parameters while maintaining model efficiency?
- What is the intuition behind the smoothing techniques in statistical language modeling, and how do they help with the sparsity issue of n-gram models?
- What is the role of the Byte Pair Encoding (BPE) token learner algorithm in text preprocessing?

Group 4

- What is model adaption and what is used for? Name at least 2 types of model adaption and explain them and provide an example for each of them.
- What does a positional encoder in the Transformer Encoder do? How does it effect the self-attention mechanism?
- How does TF-IDF differ from One Hot Encoding? What are the advantages of TF-IDF over One Hot Encoding?