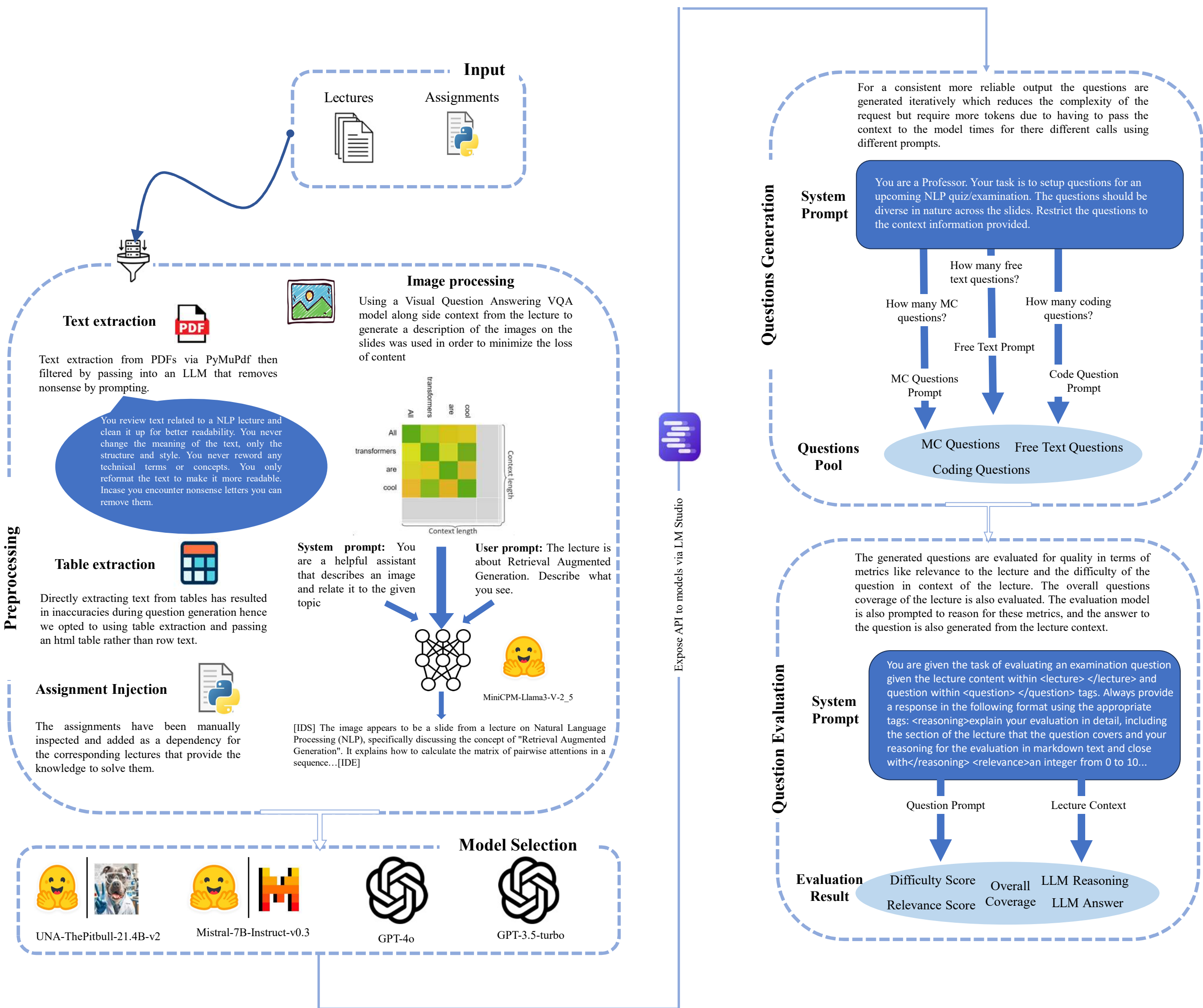# Automatic Exam Question Generation (Lecture2Exam)

Arbabha, Darius and Mohammad, Saknini, and Shaji, Shinas
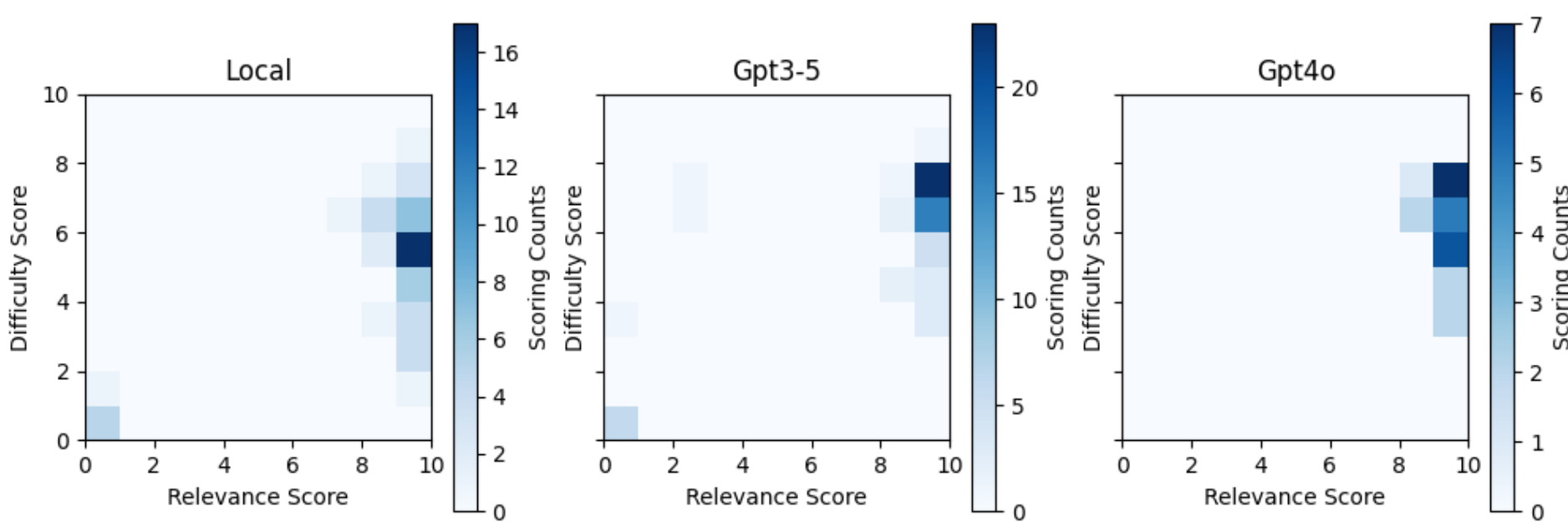
Motivation

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.
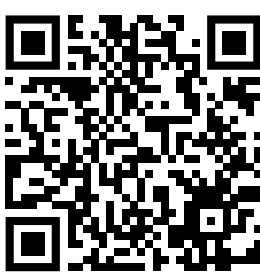
Workflow

Input

Lectures   Assignments

Preprocessing

Text extraction

Text extraction from PDFs via PyMuPdf then filtered by passing into an LLM that removes nonsense by prompting.

You review text related to a NLP lecture and clean it up for better readability. You never change the meaning of the text, only the structure and style. You never reword any technical terms or concepts. You only reformat the text to make it more readable. Incase you encounter nonsense letters you can remove them.

Image processing

Using a Visual Question Answering VQA model along side context from the lecture to generate a description of the images on the slides was used in order to minimize the loss of content

System prompt: You are a helpful assistant that describes an image and relate it to the given topic

User prompt: The lecture is about Retrieval Augmented Generation. Describe what you see.

MiniCPM-Llama3-V-2_5

[IDS] The image appears to be a slide from a lecture on Natural Language Processing (NLP), specifically discussing the concept of "Retrieval Augmented Generation". It explains how to calculate the matrix of pairwise attentions in a sequence...[IDE]

Table extraction

Directly extracting text from tables has resulted in inaccuracies during question generation hence we opted to using table extraction and passing an html table rather than row text.

Assignment Injection

The assignments have been manually inspected and added as a dependency for the corresponding lectures that provide the knowledge to solve them.

Model Selection

UNA-ThePitbull-21.4B-v2   Mistral-7B-Instruct-v0.3   GPT-4o   GPT-3.5-turbo

Expose API to models via LM Studio

Questions Generation

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 different prompts.

System Prompt

You are a Professor. Your task is to setup questions for an upcoming NLP quiz/examination. The questions should be diverse in nature across the slides. Restrict the questions to the context information provided.

How many free text questions?

How many MC questions?

How many coding questions?

Free Text Prompt

MC Questions Prompt

Code Question Prompt

Questions Pool

MC Questions   Free Text Questions   Coding Questions

Question Evaluation

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.

System Prompt

You are given the task of evaluating an examination question given the lecture content within <lecture> </lecture> and question within <question> </question> tags. Always provide a response in the following format using the appropriate tags: <reasoning>explain your evaluation in detail, including the section of the lecture that the question covers and your reasoning for the evaluation in markdown text and close with</reasoning> <relevance>an integer from 0 to 10...

Question Prompt   Lecture Context

Evaluation Result

Difficulty Score   Overall Coverage   LLM Reasoning

Relevance Score   LLM Answer

Results

Heatmap of the self-evaluated difficulty and relevance scores of generated questions

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?

Poll	Human	GPT3.5	GPT4o	Local
"Human" votes				

Contact

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

Github Repository

Acknowledgement

b-it Bonn-Aachen International Center for Information Technology

Hochschule Bonn-Rhein-Sieg
University of Applied Sciences