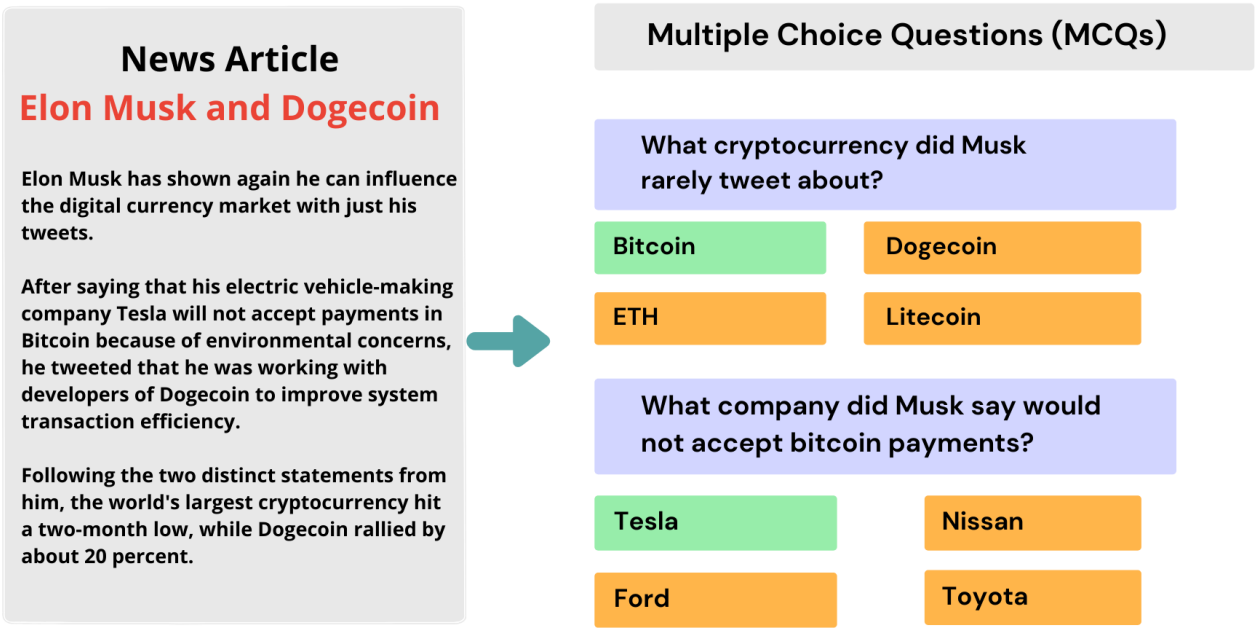
Questgen

Complete guide to generating Multiple Choice Questions automatically using AI

Hardik Kamboj · Dec 26, 2021 · 4 min read

Automating the processing of MCQ generation using state-of-the-art Natural Language Processing techniques.



MCQ's are the most commonly used assessment type and the task to create these MCQ assessments is usually cumbersome and requires a subject matter expert who reads through the content and generates appropriate questions and corresponding wrong answers (distractors).

Using advanced Natural Language Processing techniques, we will see how we can automate the process of MCQ Generation.

How would a teacher create MCQs?

Before moving on to the technical implementation, let's go through how a question author (teacher, tutor, etc) would approach this problem. This will give us some idea on how to structure out the steps required for the solution.

Step 1: Identify key Sentences/Concepts

News Article

Elon Musk and Dogecoin

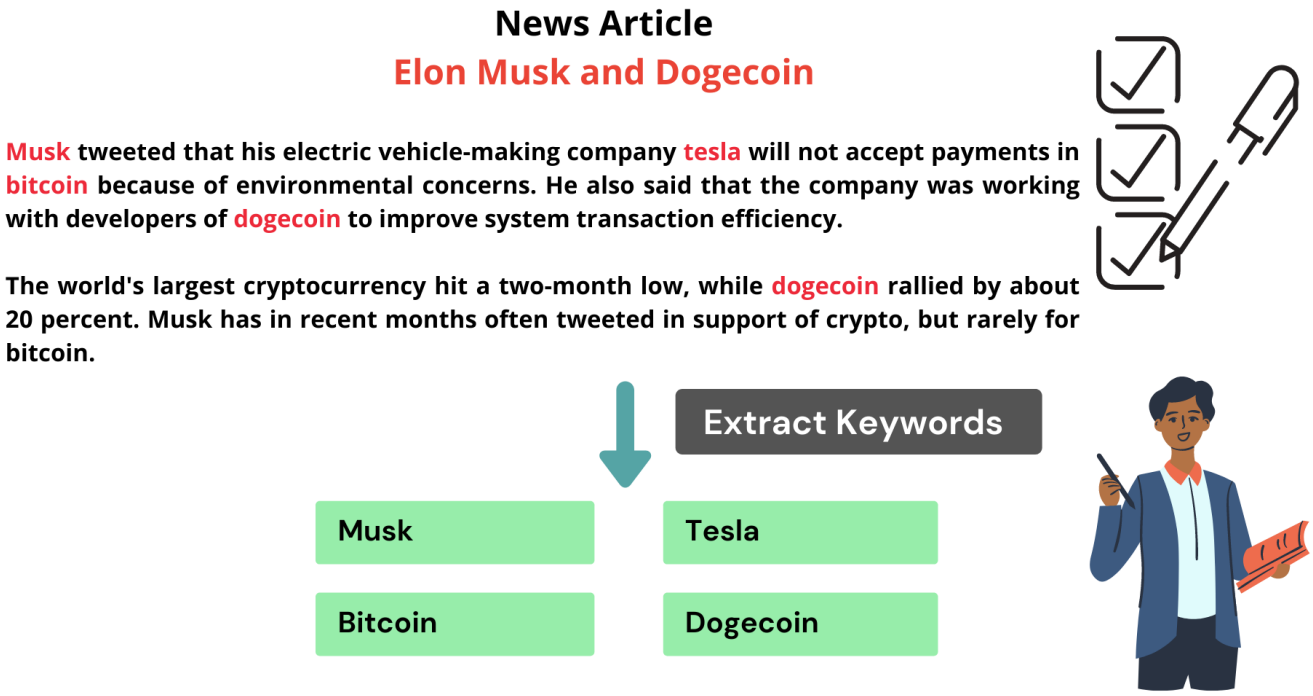
Elon Musk has shown again he can influence the digital currency market with just his tweets. After saying that his electric vehicle-making company Tesla will not accept payments in Bitcoin because of environmental concerns, he tweeted that he was working with developers of Dogecoin to improve system transaction efficiency.

Following the two distinct statements from him, the world's largest cryptocurrency hit a two-month low, while Dogecoin rallied by about 20 percent. The SpaceX CEO has in recent months often tweeted in support of Dogecoin, but rarely for Bitcoin. In a recent tweet, Musk put out a statement from Tesla that it was "concerned" about the rapidly increasing use of fossil fuels for Bitcoin mining and transaction, and hence was suspending vehicle purchases using the cryptocurrency.

A day later he again tweeted saying, "To be clear, I strongly believe in crypto, but it can't drive a massive increase in fossil fuel use, especially coal". It triggered a downward spiral for Bitcoin value but the cryptocurrency has stabilized since.

The first step would be to go through the sentences and figure out the key sentences/concepts being discussed in the given text. This would help us to filter out the sentences which are less important.

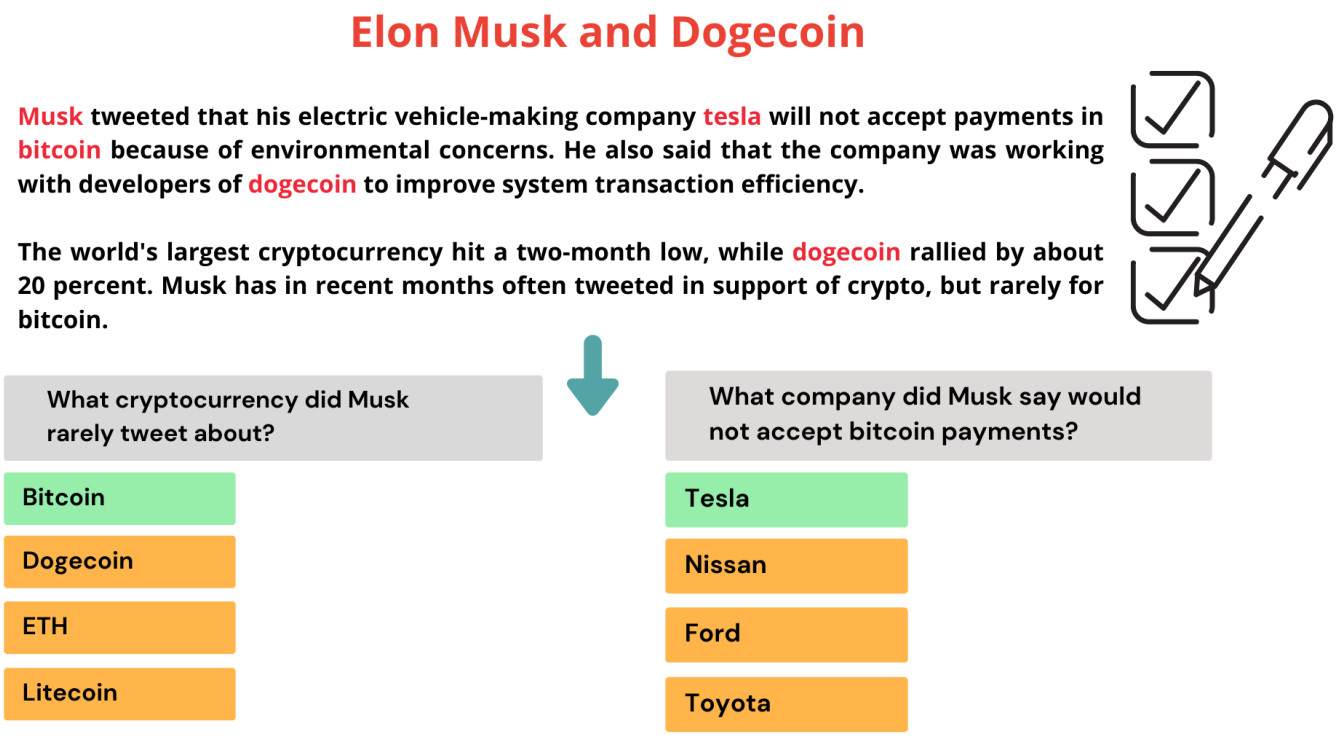
Step 2: Identify keywords from sentences



After filtering out the sentences, we need to extract the keywords or keyphrases. Keywords or keyphrases are the important concepts in the sentences which will serve as the correct answers, on the basics of which MCQs will be generated.

In the above text, Musk, Tesla, Bitcoin, Dogecoin, etc are the extracted keywords.

Step 3: Form Multiple Choice Questions



With the knowledge of the given text, the tutor/teacher can then form the multiple choice question. These questions are formed in such a way that the individual keywords or keyphrases are the answers to them.

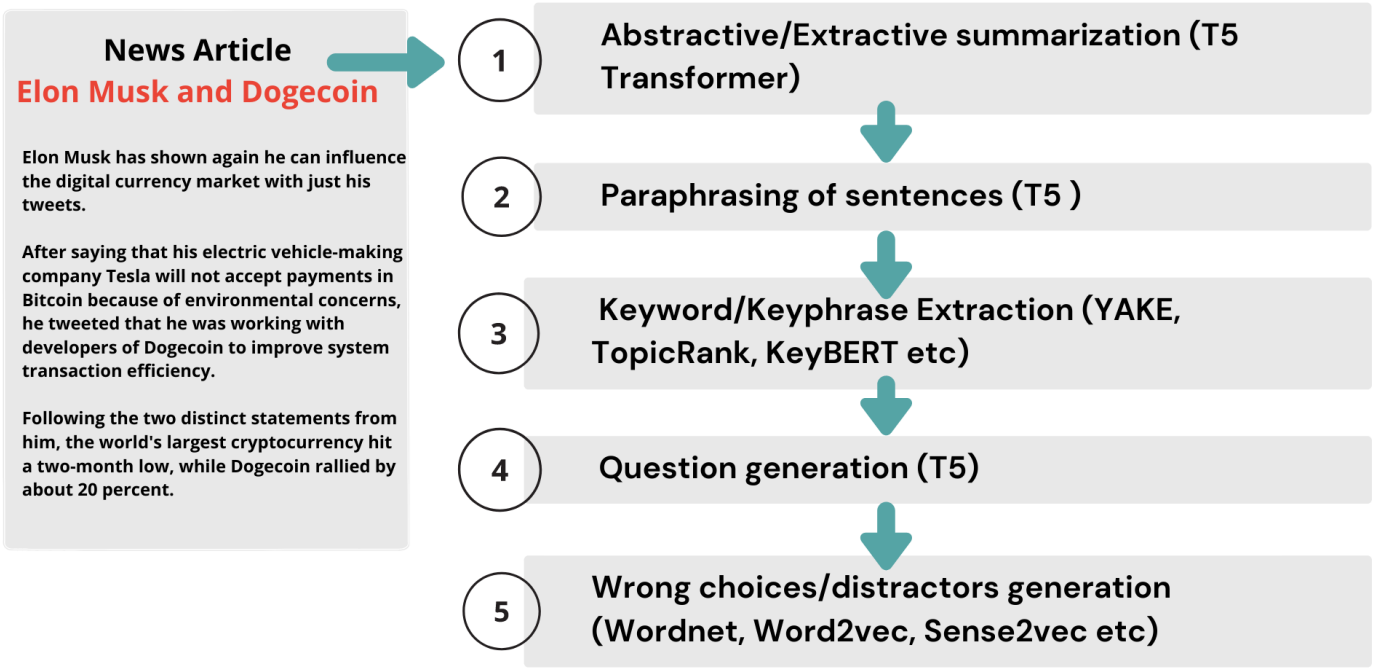
Along with the question and the correct answer, the tutor/teacher will also create distractors (wrong answer choices). The rule of thumb is

also create distractors (wrong answer choices). The rule of thumb is that the distractors should be similar to the correct answer but not obvious enough to give away the correct answer. They are meant to confuse the quiz taker.

Posing it as an NLP Problem

Having gone through the steps a teacher would intuitively use to generate the MCQs, our task now is to pose this as a Natural Language Processing problem and attempt to solve it.

The overview of this would look something like this ->



The process will contain 5 steps in total -

1. Abstractive/Extractive Summarization.
2. Paraphrasing of sentences.
3. Keyword/Keyphrase Extraction.
4. Question Generation.

5. Wrong choices/distractors generation.

Let's go through these step by step.

Step 1: Abstractive/Extractive Summarization

Summarization refers to the task of creating a short summary of the whole text. Summarization can be done in two ways, abstractive summarization, and extractive summarization. While extractive summarization extracts words and word phrases from the original text to create a summary, abstractive summarization learns an internal language representation to generate more human-like summaries, paraphrasing the intent of the original text.

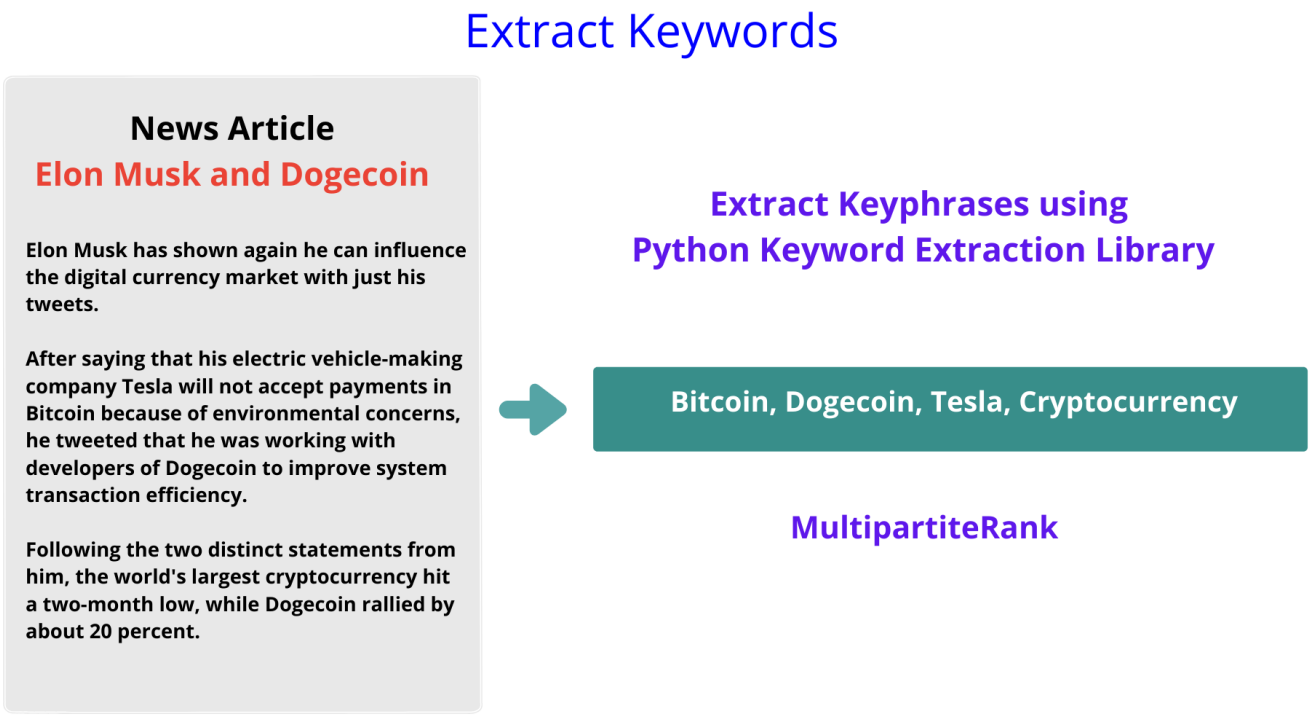
We can leverage these T5 Models for summarization.

Step 2: Paraphrasing of sentences

We would need to paraphrase the result if we are using extractive summarization in the first step. Although, if we had used abstractive summarization earlier we can skip this step as it has already been paraphrased during summarization.

We can choose amongst these paraphrases available in the Hugging Face NLP library.

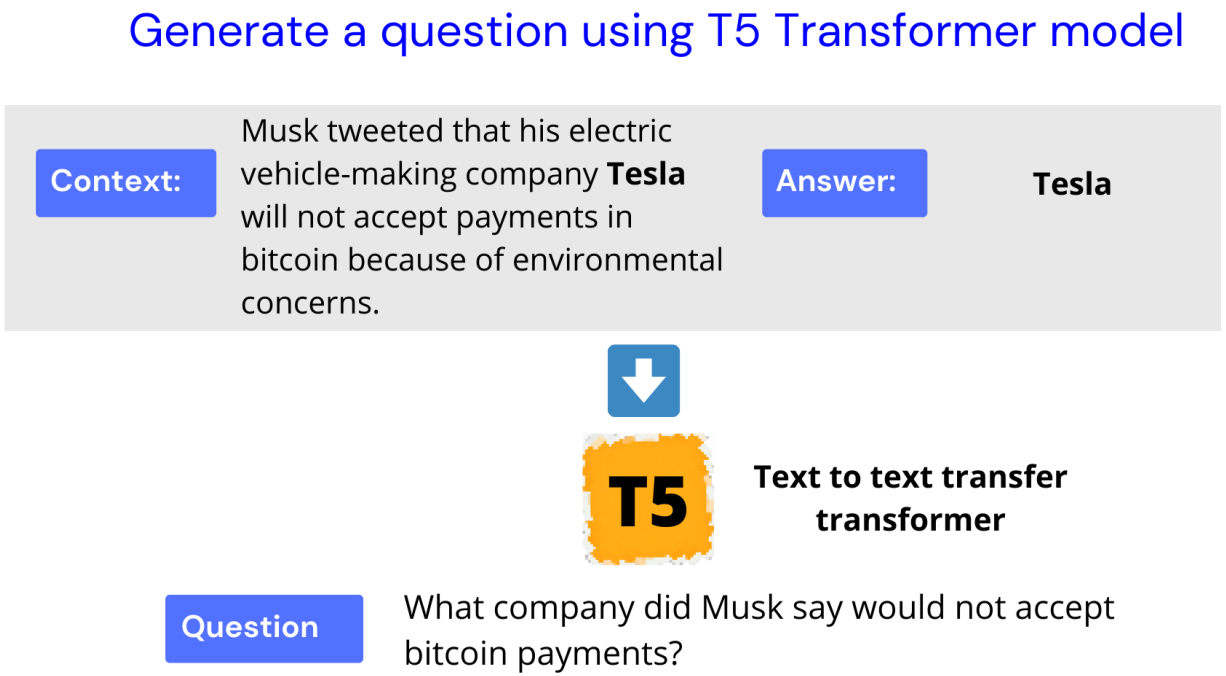
Step 3: Keyword/Keyphrase extraction



Given the paraphrased sentence, next, we would extract the keywords. These keywords would serve as the basis for questions that will be generated in the next step.

Multipartite rank is one useful algorithm, used in keyword extraction. We can leverage the open-source library python-keyphrase-extraction (pk3) for this task.

Step 4: Question Generation



We have the keywords and the text, which will now be used to generate questions related to each keyword

questions related to each keyword.

We can use the T5 Transformer model that is explicitly trained to take some context and a relevant keyword and generate an appropriate question as shown above.

Step 5: Getting distractors

We have the question and also the right answer for it. To complete the generation of MCQ, we now need to create distractors which are the wrong choices for the MCQ answer.

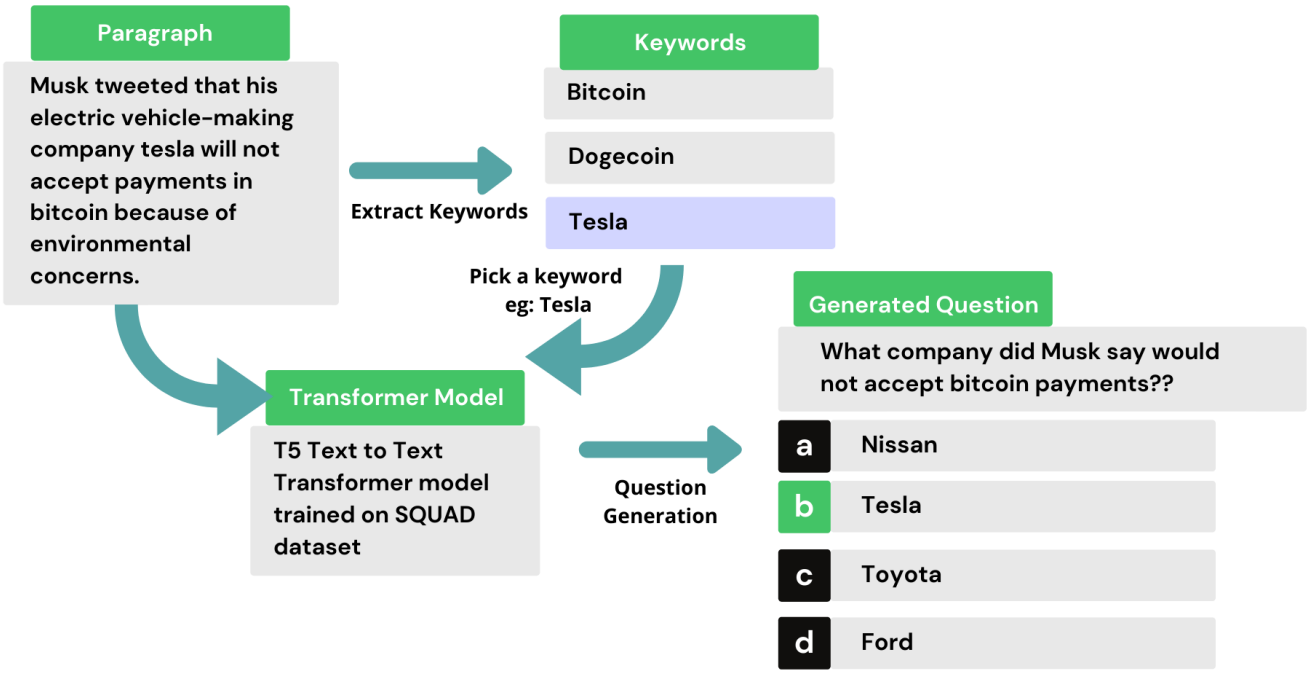
For example, if there is a question - What is the capital of India, for which the right answer is New Delhi, then the distractors can be other metropolitan cities like Mumbai, Kolkata, Chennai, etc.

We can use word vector algorithms and their variants like sense2vec to find distractors. We can also use advanced prompts to OpenAI's GPT-3 language model and generate appropriate distractors.

• • •

OVERVIEW

Overview: Generate MCQs from any content.



The diagram above shows different processes that we had used to generate Multiple Choice Questions from a given text.

artificial intelligence ev natural language processing in edtech el saas solve a real-world problem.

quiz maker

This is how we can create Multiple Choice Questions (MCQs) automatically using AI.

You may also want to read

Questgen: An Advanced AI powered question generator that is easy to use.

artificial intelligence

Ramsri Golla · Oct 24, 2021 · 2 min read