Form 1: Project Information Form

1. Team No: 10

2. Project Title: SmartQuiz: Empowering Education with Dynamic and Ethical MCQ Generation

3.Team Details:

Sl. No	Hall ticket Number	Name
1	20EG105314	Garlapati Akshay Reddy
2	20EG105325	Lakshmishetty Nikhil Shetty
3	20EG105343	Ramankol Srikanth

4.Problem Statement:
Current manual methods for creating multiple-choice questions (MCQs) are time-consuming and lack scalability, particularly in the absence of specialized datasets, such as in computer science. This project addresses the need for an efficient and automated MCQ generation system, leveraging advanced NLP techniques. The goal is to overcome limitations in diversity, efficiency, and relevance, offering a streamlined solution applicable across various subjects.

5.Source of Project:

A. Hadifar, S. K. Bitew, J. Deleu, C. Develder and T. Demeester, "EduQG: A Multi-Format Multiple-Choice Dataset for the Educational Domain," in IEEE Access, vol. 11, pp. 20885-20896, 2023, doi: 10.1109/ACCESS.2023.3248790.

6.FinalOutcome:
We will develop a web interface to generate MCQS from text and create a dataset on computer science subjects.

7.What are parameters consider for project evaluation
Text Summarization Quality, Keyword Extraction Precision , Diversity and Relevance of Distractors

8.Development Environment:

Visual Studio Code (VSCode) /Anaconda

Signature Team Members

Signature Supervisor

1

2

3

Form 3: Methodology

1. Team No: 10

2. Project Title: SmartQuiz: Empowering Education with Dynamic and Ethical MCQ Generation

3. Proposed Method

The proposed method for dynamic multiple-choice question (MCQ) generation involves a systematic approach. First, the input text undergoes summarization using BERTSUM. Keywords are then extracted using Python Keyphrase Extraction (PKE), and FlashText is applied for mapping these keywords back to their respective sentences. Subsequently, the pre-trained T5 model is fine-tuned with a specialized computer science dataset. The fine-tuned model is ready for dynamic question generation based on the mapped sentences. Sense2Vec is employed for generating contextually relevant distractors. The entire process is integrated into a user-friendly web interface, allowing seamless text input and real-time MCQ generation.

4. Proposed Method illustration

- **Step 1:** Input Text Submission:

  Users submit educational content in text format through a user-friendly web interface.

- **Step 2**: Text Summarization:

  The input text undergoes summarization using BERTSUM, condensing it to capture the essential information.

- **Step 3:** Keyword Extraction:

  Python Keyphrase Extraction (PKE) is applied to extract important keywords from the summarized text.

- **Step 4:** Keyword Mapping:

  FlashText is utilized to map the extracted keywords back to their respective sentences, ensuring contextual relevance.

- **Step 5:** Fine-tuning T5:

  The pre-trained T5 model is fine-tuned with a diverse dataset supporting the model without making it overly specific.

- **Step 6:** Question Generation:

  The fine-tuned T5 model dynamically generates multiple-choice questions based on the mapped sentences and keywords.

- **Step 7:** Distractor Generation:

  Sense2Vec is employed to generate contextually relevant distractors for the multiple-choice question options.

- **Step 8:** Web Interface Integration:

  The entire question generation pipeline is seamlessly integrated into the backend of the user-friendly web interface.

- **Step 9:** User Interaction:

  Users can interact with the web interface by submitting text, and real-time results are generated with multiple-choice questions.

5. Parameter Formulas

- **F1 Score**:
  Formula : *F*1= Precision+Recall / 2×Precision×Recall

- **ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation - Longest Common Subsequence)** :
  Formula:

  ROUGE − L = Total Words in Reference Summary / Longest Common Subsequence (LCS)

- **Maximal Marginal Relevance (MMR) for Distractor Generation:**

  For distractor generation, it aims to select distractors that balance relevance to the correct answer and diversity from already selected distractors.

  Formula:

  MMR= $\arg\max_{d_i \in D \setminus S} \left[\lambda \times \mathrm{Sim}(d_i, Q) - (1-\lambda) \times \max_{d_j \in S} \mathrm{Sim}(d_i, d_j)\right]$