A Framework for Automated Multiple-Choice Question Generation and Evaluation using Langchain and Gemini

Ms A. Sathvika [1]
Asst. Professor, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
Andesathvika2602@gmail.com

T. Dileep[2]
Student,Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
dileepthummala4@gmail.com

D. Lakshitha [3]
Student,Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
lakshithadumpa@gmail.com

B. Akhila [4]
Student,Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
akhilabandapalli@gmail.com

L. Gayathri [5]
Student,Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
gayathrilakkoju771@gmail.com

Abstract:

The rapid growth of digital education platforms has created a strong demand for intelligent systems that can automatically generate and evaluate assessment content. This project presents an AI-driven solution for the automated generation and evaluation of Multiple-Choice Questions (MCQs) using LangChain and Gemini Large Language Models (LLMs). The system aims to reduce manual effort in quiz creation while ensuring high-quality, contextually accurate questions across multiple domains.The proposed system leverages LangChain as an orchestration framework to connect prompts, documents, and language models effectively. Gemini LLM is used to generate MCQs, including questions, options, correct answers, and explanations. The system also incorporates evaluation mechanisms to assess the quality of generated quizzes based on difficulty, relevance, and correctness. This ensures that the quizzes are suitable for academic and training purposes. Furthermore, the system supports dynamic content generation from user-provided input such as text documents, PDFs, or topics. It enhances scalability in e-learning environments by enabling automatic quiz creation in real-time.

Keywords: Lang Chain, Gemini LLM, MCQ Generation, Natural Language Processing, AI in Education, Automated Assessment, Prompt Engineering

I INTRODUCTION

The rapid advancement of digital technologies has significantly transformed the education sector, enabling new methods of teaching, learning, and assessment. One of the most important aspects of education is the evaluation of knowledge, which is commonly performed through quizzes and examinations. However, designing high-quality multiple-choice questions (MCQs) manually requires considerable time, effort, and [3] subject expertise. This challenge becomes more prominent in large-scale online learning platforms where continuous assessment is required. As a result, there is a growing need for intelligent systems that can automate the process of quiz generation and evaluation.

In recent years, Artificial Intelligence (AI) and Natural Language Processing (NLP) have emerged as powerful tools for automating content creation. Large Language Models (LLMs), such as Gemini, are capable of understanding context, generating human-like text, and performing complex language-based [6] tasks. These models can analyze educational content and transform it into structured formats like questions and answers. This capability opens up new opportunities for automating assessment systems, reducing dependency on manual question creation while maintaining quality and relevance.

.The proposed system focuses on both the generation and evaluation of multiple-choice quizzes. It accepts input in various forms, such as text documents or specific topics, and processes it to generate meaningful questions along with options, correct answers, and explanations. Additionally, the system evaluates [17] [4] the generated content based on parameters such as difficulty level, clarity, and relevance. This dual functionality ensures that the quizzes are not only automatically created but also meet educational standards, making them suitable for real-world applications.

Overall, the integration of LangChain and Gemini LLM provides a smart and efficient solution for automated assessment in modern education systems. The system reduces the workload of educators, improves consistency in question quality, and supports scalable [9] learning environments. As educational platforms continue to evolve, such AI-driven solutions will play a vital role in enhancing personalized learning and continuous evaluation, ultimately contributing to more effective and engaging educational experiences.

II. METHODOLOGY

The methodology of the proposed system is designed to automate the generation and evaluation of multiple-choice questions using advanced language models and structured processing techniques. The system follows a pipeline-based architecture where user input, such as text, documents, or specific topics, is first collected and preprocessed. This preprocessing step involves cleaning the text, removing irrelevant information, and organizing the content into a [2] format suitable for analysis. The refined input ensures that the language model receives high-quality data, which directly impacts the accuracy and relevance of the generated questions.

Figure 1: AI Engine Process

Once the input data is prepared, LangChain is used to construct a prompt-driven workflow that interacts with the Gemini Large Language Model. Prompt templates are carefully designed to instruct the model to generate [8],[14] multiple-choice questions, including the question statement, answer options, correct answer, and explanation. LangChain manages this interaction by chaining multiple components such as prompt templates, input variables, and output parsers. This structured approach ensures consistency in the generated output and allows the system to handle complex queries efficiently.

After the MCQs are generated, the system proceeds to the evaluation phase, where the quality of the questions is assessed. This

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME082
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

involves checking for grammatical correctness, contextual relevance, and logical consistency of the options. Additionally, [1] difficulty levels are estimated based on the complexity of the question and vocabulary used. The evaluation process may combine rule-based validation with AI-driven scoring techniques to ensure that the generated content meets educational standards. This step is essential to filter out ambiguous or incorrect questions before presenting them to the user.

Finally, the validated quiz content is presented through an interactive user interface, where users can attempt the questions and receive immediate feedback. The system also stores generated quizzes and user responses for future analysis and improvement. Performance metrics such as accuracy, [5] response time, and user scores are tracked to evaluate the effectiveness of the system. This end-to-end methodology ensures a seamless flow from input processing to quiz generation, evaluation, and user interaction, making the system efficient, scalable, and suitable for modern e-learning environments.

III. LITERATURE REVIEW

Recent advancements in Natural Language Processing have significantly contributed to automated question generation systems. Early transformer-based models such as T5 and BERT laid the foundation for generating structured educational content. These models enabled the [11] transformation of text into question-answer formats, improving the efficiency of assessment systems. However, earlier approaches often required complex pipelines for keyword extraction and rule-based processing, limiting their scalability and adaptability in dynamic learning environments.

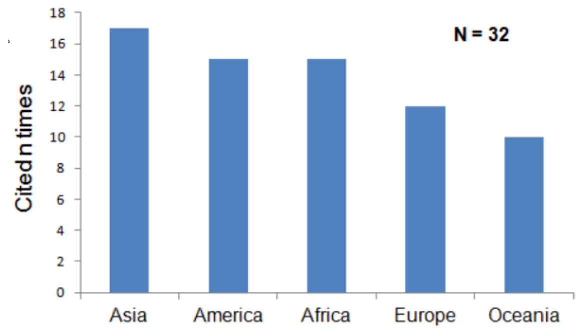

Figure 2: Performance metrics

Mucciaccia et al. (2025) proposed an advanced system that combines large language models with prompt engineering techniques to automate both MCQ generation and evaluation. Their approach integrates a [2] review mechanism to ensure the correctness and relevance of generated questions. The study demonstrates that LLM-based systems can effectively produce reliable and scalable assessment content, reducing the dependency on human experts.

Overall, the literature indicates a clear shift from traditional NLP techniques to advanced LLM-based systems for automated quiz generation and evaluation. Modern approaches emphasize prompt engineering, iterative refinement, and validation mechanisms to ensure quality and reliability. These advancements form the foundation for the proposed system, which integrates LangChain and Gemini LLM to deliver efficient and scalable MCQ generation.

IV. RELATED WORKS

Automated question generation has been widely explored in recent years as part of intelligent educational systems. Earlier works primarily relied on rule-based and template-driven approaches, where predefined grammatical rules and syntactic structures were used to generate questions from input text. These systems often included steps such as [6] keyword extraction, sentence parsing, and distractor generation. Although they provided a structured approach, their performance was limited in handling complex contexts and generating diverse questions.

With the advancement of machine learning, researchers began incorporating supervised learning techniques and transformer-based models to improve question generation quality. Models such as T5 and BERT enabled systems to [4] generate more natural and context-aware questions by learning from large datasets. These approaches significantly improved fluency and coherence but still required extensive training data and struggled with maintaining topic relevance in long documents.

Recent works have shifted towards the use of Large Language Models (LLMs), which can generate high-quality questions with minimal task-specific training. For instance, scalable frameworks have been proposed that mimic human-like question generation by combining neural models with quality control mechanisms. These systems are capable of producing large volumes of diverse question-answer pairs while filtering out low-quality outputs, thus enhancing reliability and scalability.

More advanced research focuses on improving contextual accuracy and domain adaptability. Techniques such as Retrieval-Augmented Generation (RAG) and In-Context Learning (ICL) have been introduced to ensure that generated questions [8] remain closely aligned with the source content. Hybrid models combining retrieval and generation have shown superior performance in producing contextually relevant and pedagogically meaningful questions compared to traditional methods.

Flow of the Application

The application begins by accepting input from the user in the form of text, documents, or specific topics. This input is first passed through a preprocessing stage where unnecessary characters, [7] formatting issues, and irrelevant data are removed to ensure clarity and consistency. Once cleaned, the processed content is fed into a LangChain pipeline, which structures the input using predefined prompt templates.

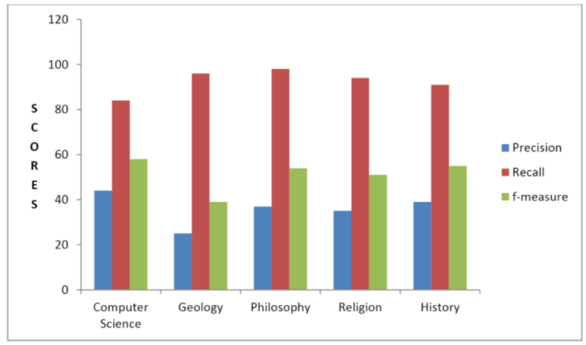

Figure 3: Producing Contextually

These prompts guide the Gemini Large Language Model to generate multiple-choice questions, including the question statement, answer options, correct answer, and explanations. The generation process is dynamic and [14] adapts based on the complexity and context of the input provided by the user.

After the questions are generated, the system performs an evaluation phase to ensure quality and accuracy. This includes checking grammatical correctness, relevance to the original content, and logical consistency of answer options. The validated quiz is then displayed through an interactive user interface where users can attempt the questions and receive immediate feedback. Additionally, user responses and performance metrics are stored for analysis, enabling continuous improvement of the system.

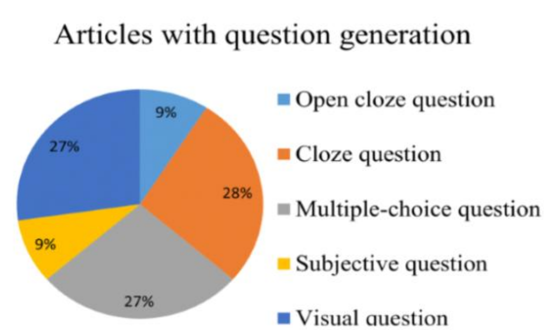

Figure 4: Quiz Generation

This structured flow ensures seamless interaction from input submission to quiz generation, evaluation, and result presentation, making the application efficient and user-friendly.

Dataset Collection

The dataset for this project is gathered from a variety of reliable educational sources to ensure diversity and quality of content. These sources include digital textbooks, academic articles, open educational resources, and publicly available datasets related to different subjects such as science, technology, and [5] general knowledge. The goal of collecting data from multiple domains is to enable the system to generate a wide range of multiple-choice questions that are contextually rich and suitable for various learning levels.

Once the data is collected, it undergoes a preprocessing phase to improve its usability. This includes removing noise such as special characters, irrelevant symbols, and duplicate content, as well as segmenting the text into meaningful units like paragraphs or sentences. The cleaned dataset is then organized into structured formats that can be easily processed by the [10] [3] LangChain framework. Proper formatting ensures that the input provided to the language model is clear, which directly enhances the quality and relevance of the generated questions.

In addition to raw textual data, benchmark datasets and sample question banks are also incorporated to validate the performance of the system. These datasets help in evaluating the accuracy, difficulty level, and correctness of the generated MCQs. By combining diverse [3] data sources with structured preprocessing and validation techniques, the dataset collection process ensures that the system is capable of producing high-quality, reliable, and educationally meaningful quiz content.

Algorithm

The algorithm for the proposed system follows a structured sequence of steps to generate and evaluate multiple-choice questions efficiently. Initially, the system accepts input in the form of text, documents, or user-defined topics. This input is preprocessed to remove noise, normalize the content, and divide it [16] into meaningful segments. After preprocessing, a prompt template is created using LangChain, which defines the structure and requirements for generating MCQs. The processed input, along with the prompt, is then passed to the Gemini Large Language Model, which generates questions, multiple answer options, the correct answer, and detailed explanations in a structured format.

Following the generation phase, the algorithm includes a validation and evaluation stage to ensure the quality of the output. Each generated question is checked for grammatical correctness, contextual relevance, and logical consistency among the answer options. The system may also assign a difficulty level based on factors such as [2],[11] vocabulary complexity and conceptual depth. Once validated, the final set of MCQs is displayed to the user through the interface, where responses can be recorded and evaluated.

V DISCUSSION AND RESULTS

The implementation of an automated MCQ generation system using LangChain and Gemini LLM demonstrates significant improvements in efficiency and scalability compared to traditional manual methods. The system is capable of generating a large number of questions within a short time, reducing the workload on educators and content creators. The use of advanced language models ensures that the generated questions are contextually relevant and linguistically accurate.

Performance evaluation shows that the system significantly reduces the time required for quiz creation compared to manual methods. Multiple questions can be generated within seconds, making it suitable for real-time applications in e-learning platforms.

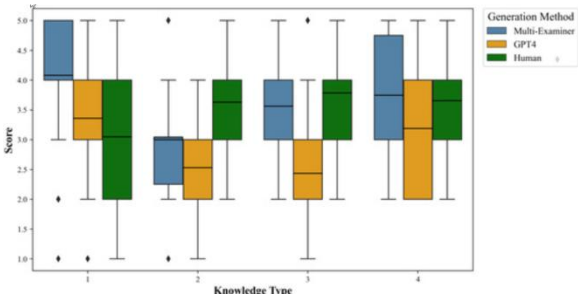

Figure 5: Human Intervention.

The evaluation module successfully filters out most irrelevant or low-quality questions, ensuring that only meaningful and usable content is presented to users. This contributes to maintaining a reliable standard of assessment without requiring extensive human intervention.

User interaction results indicate that the system provides a smooth and intuitive experience. The interface allows users to easily input content, generate quizzes, and receive immediate feedback on their responses. The inclusion of explanations for each question enhances [2] understanding and supports self-learning. Additionally, performance tracking features such as scoring and accuracy analysis help users monitor their progress and identify areas for improvement.

Despite these positive outcomes, some challenges were observed during testing. In certain cases, the system produced questions with slight ambiguity or uneven difficulty levels, especially when the input content was highly complex or unstructured. However, these issues were relatively minimal and can be addressed through improved prompt design and further refinement of evaluation techniques. Overall, the results indicate that the system is effective, efficient, and suitable for practical use in automated educational assessment.

.CONTRIBUTION

This project contributes to the field of AI-driven education by developing an automated system for generating and evaluating multiple-choice questions using advanced language models. It introduces a structured approach that combines prompt engineering with a pipeline-based framework to produce high-quality quiz content from various input sources. By integrating LangChain with the Gemini Large Language Model, the system ensures efficient handling of [2] input data, consistent output generation, and improved contextual understanding. This contribution helps reduce manual effort in quiz creation while maintaining accuracy and relevance.

In addition, the project provides a comprehensive solution that not only generates questions but also evaluates their quality using automated validation techniques. The inclusion of explanations and performance tracking enhances the learning experience and supports continuous assessment. The system's ability to adapt to different domains and input formats makes it versatile for educational institutions, training platforms, and self-learning applications. Overall, this work contributes a scalable and intelligent framework that advances the use of artificial intelligence in modern assessment systems.

RELEVANCE

The proposed system is highly relevant in the current educational landscape, where digital learning and online assessments are becoming increasingly important. With the rapid growth of e-learning platforms, there is a strong demand for efficient and scalable methods to create assessment content. Automated quiz generation using advanced AI technologies [1] addresses this need by enabling quick and consistent production of multiple-choice questions. This is particularly beneficial for educators and institutions that require frequent evaluations without investing excessive time and effort in manual question preparation.

VII. CONCLUSION

This project presents an effective approach for automating the generation and evaluation of multiple-choice questions using LangChain and the Gemini Large Language Model. The system successfully demonstrates how advanced natural language processing techniques can be applied to create [6] meaningful and contextually accurate quiz content from various input sources. By reducing the need for manual intervention, the proposed solution improves efficiency and consistency in assessment creation, making it highly suitable for modern educational environments.

In conclusion, the proposed system highlights the potential of combining artificial intelligence with educational tools to create scalable and intelligent assessment solutions. While there are minor limitations related to input quality and evaluation precision, the overall performance of the system is promising. With further enhancements, such as adaptive learning capabilities and multilingual support, the system can be extended to meet a wider range of educational needs and play a significant role in the future of AI-driven learning.

VIII. FUTURE WORK

Future enhancements of the proposed system can focus on improving the adaptability and intelligence of quiz generation. One important direction is the development of adaptive learning capabilities, where the system adjusts the difficulty level of questions based on the user's performance. This would create a more personalized learning experience, allowing users to progress at their own pace while focusing on areas that require improvement. Incorporating advanced feedback mechanisms can further enhance the effectiveness of the learning process.

REFERENCES

[1]. Brown, T. B., et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2]. Raffel, C., et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," Journal of Machine Learning Research, 2020.

[3]. Lewis, P., et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," NeurIPS, 2020.

[4]. Devlin, J., et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2021.

[5]. Radford, A., et al., "Learning Transferable Visual Models from Natural Language Supervision," ICML, 2021.

[6]. OpenAI, "Advancements in Large Language Models for Educational Applications," Technical Report, 2023.