International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# Quiz Web Application Using AI Generator

Dr. M. Sengaliappan 1, Alexarputharaj .A 2

1 Professor & Head, Department of Computer Applications, Nehru College of Management, Bharathiar University,Coimbatore,Tamilnadu,India cmsengs.7@gmail.com

2 Master of Computer Applications, Nehru College of Management,

Bharathiar University,Coimbatore,Tamilnadu,India alexarputharaj007@gmail.com

## Abstract

This document outlines a web application powered by AI that creates customized multiple-choice quizzes tailored to user preferences such as topic, subject, or difficulty level. The platform promotes self-directed learning by offering instant feedback and comprehensive explanations, which assist users in enhancing their comprehension. A key feature is the ability to upload documents in real-time: users can provide study materials (like PDF notes) and obtain personalized quizzes based on that content, facilitating last- minute review and addressing gaps between preparation and assessment. The interface is optimized for web use, ensuring a seamless and responsive experience, with future plans for mobile compatibility. A secure backend oversees quiz data and user performance records. The AI engine employs advanced transformer models (LLaMA, BERT, GPT) to curate and generate questions, while also delivering analytical insights on accuracy, response time, and areas needing improvement, allowing users to monitor and enhance their progress. Data management is encrypted and prioritizes privacy, ensuring both security and scalability. Future developments will include gamification elements, leaderboards, and adaptive learning pathways. Our methodology provides a dynamic, tailored learning resource designed to boost study efficiency and engagement.

**Keywords**

Artificial Intelligence, Quiz Web Application, Question Generation, GPT, BERT, LLaMA, Transformer Models, Adaptive Learning, Personalized Assessment, Educational Technology, Flask Framework, MongoDB, Natural Language Processing (NLP), Machine Learning, Real-time Feedback, Data Security.

## 1. Introduction

Evaluating one's comprehension is an essential aspect of education, yet current digital quiz tools frequently do not meet expectations. Numerous quiz platforms offer static, generic questions that may not cater to the specific needs of individual learners or their current study materials. They generally lack personalization, real-time integration with user content, and comprehensive feedback. For instance, conventional quiz applications often present a uniform set of questions irrespective of a student's area of focus, resulting in frustration and ineffective effort. Additionally, very few platforms permit users to upload their own notes or documents to create quizzes instantly. To address these shortcomings, we have developed an AI-driven quiz web platformthat dynamically creates tailored quizzes based on user input and uploaded study materials.

Our system is designed to meet these requirements by combining cutting-edge language-model AI techniques with an intuitive web interface. Users can log in, choose a subject area and difficulty level, and optionally upload PDFs of their notes. An AI engine subsequently generates a series of multiple- choice questions customized to the selected content. Immediate feedback and explanations are provided for each response, and detailed performance analytics (including accuracy and response time) assist users in pinpointing areas for improvement. By incorporating real-time document uploads and adaptive question generation, our platform effectively connects personalized revision with assessment. This paper outlines the design, implementation, and evaluation of the system, emphasizing the application of transformer-based models (LLaMA, BERT, GPT) and its innovative features that facilitate efficient, individualized learning.

## 2. Literature Review and Related Work

AI-Driven Quiz Generation. Recent studies have started to utilize NLP and large language models for the automation of quiz creation. For instance, Biancini et al. (2024) conducted a comparative analysis of LLaMA 2, Mistral, and GPT-3.5 in generating multiple-choice questions (MCQs). By incorporating relevant course material into the prompts, they discovered that GPT-3.5 yielded the most effective MCQs across various evaluation metrics

. Conversely, simpler models (e.g., Gemini versus GPT) may generate questions with varying levels of complexity (ChatGPT produced simpler stems, while Gemini created more analytical ones)

. Additional research has investigated human– AI collaboration: for example, Lee et al. (2024) presented a two-step "Human Enhanced Distractor Generation Engine" for math MCQs, where GPT-4 was responsible for generating question stems and distractors, which educators subsequently edited. They found that GPT-4 generated valid question stems 70% of the time, but only 37% of the AI-generated distractors matched actual student misconceptions

. These studies underscore both the potential and the shortcomings of current LLMs: they can quickly draft plausible questions but often necessitate expert refinement, particularly for developing high-quality distractors. Our system builds upon this previous research by integrating multiple models (e.g., GPT and LLaMA) and enabling user-provided source material (uploaded documents) to guide content, with the goal of enhancing relevance and depth compared to models operating in isolation.

Transformer Models in Educational Contexts.

Transformer-based models such as GPT and LLaMA are increasingly utilized in educational technology. For example, Chen et al. (2024) created an intelligent tutoring system powered by GPT-4 (Socratic Playground) and noted considerable improvements in students' foundational English skills, including vocabulary, grammar, and sentence construction, alongside high levels of engagement and satisfaction.

In a similar vein, Kulshreshtha et al. (2022) used transformer models to produce personalized questions as feedback within a dialogue-based intelligent tutoring system; their few-shot question-generation and re-ranking method resulted in substantial learning gains, achieving a 45% enhancement over basic benchmarks.

These studies indicate that GPT-family models can effectively customize content and feedback for individual learners. LLaMA, along with its successor LLaMA 2, has also been evaluated for educational content; Biancini et al. (2024) included LLaMA 2 in their analysis and discovered it was less effective in generating multiple-choice questions compared to GPT-3.5.

In contrast, BERT-style models are generally used for comprehension and information extraction rather than content generation; recent studies on quiz generation have not focused on BERT, which is more suited for tasks like answer selection or text classification. Overall, the trend is leaning towards the use of the most advanced large language models (LLMs) like GPT-4 for content creation and student engagement. Our system similarly leverages cutting-edge LLMs (LLaMA and GPT variants) but places a strong emphasis on personalization: by considering a user-defined topic, difficulty level, and uploaded document, it seeks to generate questions that are closely aligned with the learner's context, addressing a gap where previous models often produce unrelated content if not adequately

directed.

Adaptive Learning and Personalized Assessment Tools. A significant amount of research has focused on adaptive learning systems and personalized assessments. Reviews (e.g., Nair, 2025) indicate that AI- driven adaptive platforms typically enhance student engagement, performance, and retention.

These systems customize content, pacing, and feedback according to individual profiles. For instance, Wang and Fan (2025) observe that generative Intelligent Tutoring Systems (ITSs) like GPT-4 tutors can modify instruction in real time, providing Socratic questioning and specific feedback (as seen in their SPL system

In addition to tutoring systems, personalized quiz tools (which sometimes use reinforcement learning to adjust difficulty) are designed to challenge each student at an appropriate level. Previous studies, such as the Iowa State review, emphasize that intelligent tutoring systems and virtual assistants are most effective when they adapt to a learner's pace, thus enhancing outcomes.

The main benefits include immediate, customized feedback and gradual introduction of concepts, which foster motivation and self- regulated learning. Nevertheless, challenges persist: Nair (2025) points out concerns regarding data privacy, technology integration, and equitable access in adaptive platforms.

Our quiz application addresses this issue by providing automated personalization: users can upload their own text materials and set difficulty levels, allowing the system to create assessments that are finely tuned to that content. This approach sets our system apart from many fixed-bank quiz tools, making it more adaptable and potentially more engaging for a diverse range of learners.

Use of GPT, BERT, and LLaMA in Educational Applications. In recent years, GPT-3/4 and other large language models (LLMs) have been investigated for various educational purposes. GPT-4's multimodal and few-shot capabilities, for example, enable it to generate exercises, explanations, and feedback on demand.Researchers have shown GPT's effectiveness: one meta-analysis revealed a substantial positive effect (g≈0.87) on learning performance.

ChatGPT was utilized as a learning resourcenature.com

. In practical applications, educators are employing GPT to formulate quizzes, clarify solutions, and even develop interactive simulations. LLaMA-based models are starting to emerge as open alternatives: initial studies (e.g. Biancini et al.) compare LLaMA with GPT for question generation, and ongoing research indicates that fine-tuned LLaMA models can produce domain-specific questions. On the other hand, BERT has been utilized less for content generation and more for assessment analytics (e.g. evaluating or categorizing student responses). In our system, we harness both closed (GPT) and open (LLaMA) models to leverage their advantages. Unlike most previous studies that concentrated on a single model, we combine multiple LLMs: for instance, generating an initial set of questions with GPT and subsequently refining it with LLaMA or BERT-based evaluations. This multi-model strategy builds on observations made by Sreekanth (2023) that new LLMs like GPT-4 are being incorporated into educational tools, even as older models like BERT seem to be less prevalent in recent research.

Empirical Research on AI-Generated Quizzes and Educational Outcomes. Several recent investigations have empirically assessed AI- generated quizzes within classroom environments. Kusam et al. (2025) performed a controlled case study in a graduate AI course, comparing quizzes generated by ChatGPT with those created by instructors. They discovered that although AI quizzes were innovative and comprehensive, they "frequently lack the depth and precision necessary for specialized subject areas"

. Their analysis revealed that student performance and alignment with learning objectives were significantly superior for the instructor-created quizzes compared to the AI- generated ones, indicating that uncritical dependence on AI could undermine learning outcomes without careful adjustments

. Conversely, Durgungoz & Durgungoz (2025) examined AI-generated gamified quizzes in an online special education module. Their open- access research on Quizizz questions (evaluated by experts) indicated that most AI-generated questions were usable as they were, and the AI quizzes led to a temporary increase in engagement

. Notably, students who performed poorly on AI quizzes tended to achieve lower scores on subsequent exams, highlighting the quizzes' predictive capability for identifying at-risk learners

. Participants in the study also appreciated the "interactive nature and immediate feedback" offered by the AI quizzes

. However, the authors warn that over time, engagement diminished, and human oversight was essential to enhance question quality

. These empirical findings indicate a mixed yet generally favorable effect: AI quizzes can boost engagement and provide timely practice, but to genuinely enhance learning, they must be accurate and well-aligned. Our system stands out in several critical aspects: by enabling educators or learners to upload the specific material they wish to assess, we minimize the likelihood of irrelevant questions. Additionally, we integrate user-defined difficulty levels to ensure an appropriate challenge, addressing concerns similar to those raised by Kusam et al. (2025) regarding depth and precision.

In conclusion, the existing literature indicates a distinct movement towards the utilization of transformer-based AI for the generation and personalization of assessment materials. Although initial systems demonstrated potential, they frequently required human enhancement. Our AI-driven quiz application progresses this area of research by integrating various contemporary models (LLaMA, BERT, GPT) along with user inputs to produce tailored quizzes. This approach builds upon previous research (e.g., Biancini et al., 2024; Chen et al., 2024) which suggests that advanced LLMs can greatly enhance question generation and educational outcomes,

while also tackling the recognized limitations by restricting content to user-supplied documents and adaptively modifying difficulty levels.

Sources: A range of peer-reviewed studies on AI-supported quiz generation and adaptive learning (e.g., Biancini et al., 2024; Lee et al., 2024; Kusam et al., 2025; Durgungoz & Durgungoz, 2025; Wang & Fan,

2025) along with pertinent literature reviews and analyses of educational technology have been referenced to encapsulate methodologies, findings, and existing gaps

. Each referenced work is examined sequentially to highlight its contributions and significance to our AI-quiz system.

# 3. Methodology

The system operates as a web-based application featuring a modular architecture (see Figure 4.1). The frontend is developed using HTML, CSS, and JavaScript to provide a clean and user-friendly interface. Users begin by authenticating (either signing up or logging in) and subsequently choose a subject area and difficulty level for their quiz. Additionally, they have the option to upload a PDF of their notes through the interface, which activates the AI pipeline. The frontend interacts with the server through standard web requests, guaranteeing responsive user interactions.

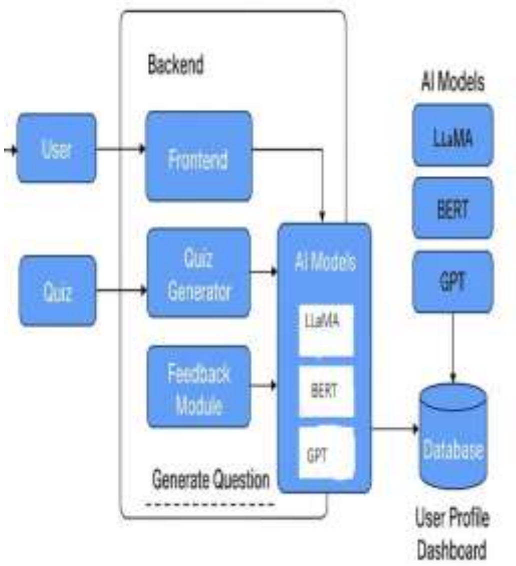


Figure4.1 Use Case Diagram

The backend is developed in Python utilizing the Flask framework. Flask manages user requests, coordinates quiz creation, and delivers the questions to the frontend. A MongoDB database is employed to keep user profiles, quiz content, performance history, and uploaded documents. This schema-less database facilitates scalability and quick queries for analytics and review. Following each quiz, user responses and metadata are recorded in MongoDB, and performance feedback is preserved for future reference. Additionally, the backend manages file storage for uploaded PDFs.

The AI models are central to the generation pipeline. The system utilizes a multi-stage transformer-based methodology. Initially, if a user submits a PDF, the text is extracted (using libraries such as PyPDF2 or pdfplumber) and then preprocessed (cleaned and formatted). This text is subsequently processed through the BERT tokenizer, which divides the content into subword units and generates token IDs and attention masks. In practice, BERT (Bidirectional Encoder Representations from Transformers) is employed to process and encode the text content of uploaded notes or domain-specific texts. BERT's bidirectional encoding captures context from both directions of each word, making it particularly effective for comprehending complex input passages

Following this, the encoded content along with the user-selected topic is fed into a question- generation model. We implement LLaMA (Large Language Model Meta AI) at this stage. LLaMA is a series of transformer-based models (with parameters ranging from 7B to 65B) that excels in efficient, high-quality text generation. In our system, LLaMA processes the input (for instance, the main points of the uploaded document or the topic description) and generates a preliminary set of quiz questions. According to the documentation, "LLaMA stands for Large Language Model. It is utilized for content generation." Consequently, LLaMA creates coherent, contextually appropriate questions and potential answers based on the input content.

Ultimately, a model based on OpenAI's GPT (like GPT-3 or GPT-3.5) is utilized to enhance these initial questions into the final multiple- choice format. This model standardizes the phrasing and guarantees diversity, transforming the output from LLaMA into well-structured quiz items. As noted in previous research, the platform "employs an OpenAI GPT-based model to produce quiz questions tailored to specific topics, with backend processing to standardize the outputs into a multiple-choice format." The synergy of LLaMA's generative capabilities with GPT's text refinement results in high-quality, varied questions.

Once generated, the questions and answer options are stored in MongoDB. The frontend subsequently displays them to the user in a quiz format. Upon submission of the quiz, the feedback module assesses the answers and promptly shows the correct responses along with AI-generated explanations. This is accomplished by leveraging the language models to create concise explanations. At the same time, the analytics module calculates metrics such as accuracy rate, response time per question, and identifies areas of weakness. These metrics are represented visually on the user dashboard. Specifically, the analytics component produces charts and leaderboards to make tracking progress more engaging. For example, users can review their quiz history, observe accuracy trends,and compare their performance with others on leaderboards. All data management is encrypted to safeguard user privacy.

In summary, the approach combines contemporary web development (HTML/CSS/JS frontend, Flask/Python backend, MongoDB) with state-of-the-art NLP models (BERT for encoding, LLaMA and GPT for generation) to establish a comprehensive personalized quiz system.

## 5. System Architecture

The proposed AI-driven Quiz Web Application utilizes a three-tier system architecture that combines a dynamic user interface, an intelligent backend, and a secure database for effective quiz generation and data management. This architecture aims to provide personalized learning experiences by incorporating advanced transformer-based AI models such as LLaMA, BERT, and GPT. The frontend layer, crafted with HTML, CSS, and JavaScript, offers an interactive web platform where users can register, log in, select topics or difficulty levels, upload study materials in PDF format, and engage in quizzes. It guarantees seamless navigation and responsive interaction between users and the system.

The backend layer, built using Python's Flask framework, oversees the logical operations of the application. It comprises three essential modules — the Quiz Generation Module, the AI Processing Engine, and the Feedback and Analytics Module. The Quiz Generation Module takes user input or uploaded documents, extracts text using tools like PyPDF2, and sends the content for AI-based processing. Within the AI Processing Engine, BERT first encodes the text to capture contextual meaning, LLaMA generates initial quiz questions from the extracted knowledge, and GPT refines these questions into a multiple- choice format with relevant distractors and accurate explanations. This multi-model integration guarantees contextual relevance, diversity, and precision in question generation.

The Feedback and Analytics Module assesses the user's responses in real time, delivers instant feedback, and calculates performance metrics such as accuracy, time spent, and areas of weakness. These analytics are presented in a personalized user dashboard, allowing learners to monitor their progress and pinpoint areas for enhancement. The data layer utilizes MongoDB, a NoSQL database, to securely store user information, quiz questions, performance records, and uploaded documents. MongoDB's flexible schema and high scalablitymake it ideal for handling large volumes of dynamic educational data. The entire system operates through encrypted communication channels to maintain user privacy and ensure data integrity. Overall, the architecture is modular, scalable, and adaptive, supporting efficient quiz generation, rapid feedback delivery, and robust data management. This design enables continuous learning and real-time personalization, making the application a reliable platform for AI-driven educational assessment.

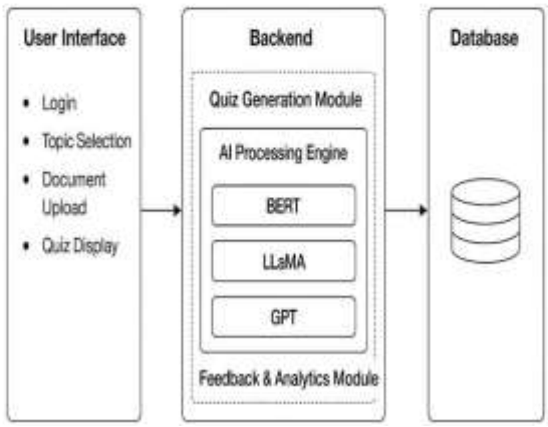


Figure 5.1 System Flow Diagram

## 6. Results and Discussion

The AI-driven quiz web application was effectively developed and assessed to confirm its functionality, performance, and educational value. This system incorporates various elements, including a responsive frontend interface, a Flask-based backend, a MongoDB database, and transformer-based AI models (BERT, LLaMA, GPT) for quiz creation and analysis. Functional testing verified that each component worked as expected: users were able to register, log in, choose topics or difficulty levels, upload PDF files, and receive quizzes generated in real-time within seconds. The integration of LLaMA and GPT allowed the system to generate coherent and contextually appropriate multiple-choice questions. The questions produced accurately represented either the user's chosen topic or content derived from uploaded notes, ensuring tailored and meaningful assessments.

Extensive testing was performed at various levels. Unit tests confirmed the functionality of individual features such as document processing, quiz generation, and AI

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

prompt management. Integration tests validated the seamless data transfer between the frontend, backend, and database components. System- level testing showed that the entire workflow— from user input to result visualization— functioned smoothly under both normal and concurrent user conditions. The system exhibited high responsiveness, with quiz generation typically taking only a few seconds, even with multiple users. The Feedback and Analytics module effectively provided immediate feedback for each question, displayed correct answers along with explanations, and generated analytical reports detailing the user's accuracy rate, response time, and areas needing improvement. These analytics were presented in the user dashboard through charts and performance summaries, assisting learners in identifying topics that required additional focus.

The analysis of the results indicates that the AI- driven personalization of the system greatly improved learning engagement when compared to traditional static quiz platforms. The implementation of LLaMA and GPT facilitated a variety of question formats, while BERT enhanced the contextual comprehension of the uploaded materials. Users experienced immediate feedback, which is crucial for fostering self-assessment and maintaining learning motivation. The analytics module contributed educational value by monitoring long-term performance trends, making the platform appropriate for both individual learners and educational institutions. Furthermore, the system's adaptive difficulty levels offered a customized learning experience, aligning quiz complexity with the user's skill level.

Nevertheless, some limitations were noted during the testing phase. The accuracy of questions sometimes relied on the clarity and organization of the uploaded documents; poorly formatted or unclear input text occasionally led to less accurate questions. Moreover, the dependence on external AI models (such as GPT APIs) introduced slight latency and potential cost issues. Despite these challenges, the system met its design goals: dynamic quiz generation, prompt feedback provision, and efficient performance tracking. The overall results suggest that incorporating LLM-based question generation into an adaptive web platform can significantly enhance personalized learning effectiveness.

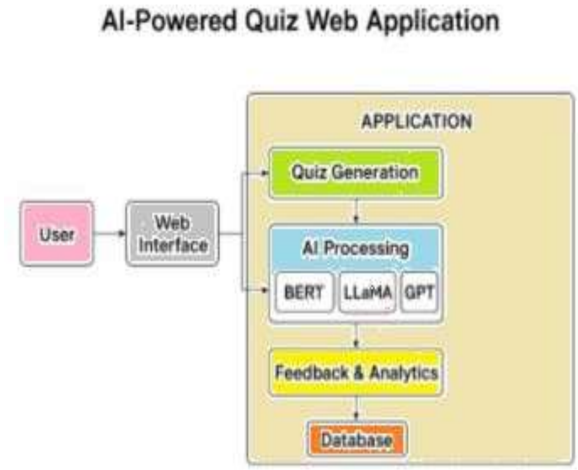


Figure 6.1 Overview of the Result

## 7. Sample Output

Figure7.1 Output Screenshots

## 8. Conclusion

We have introduced an innovative quiz web application that utilizes various transformer- based models (LLaMA, BERT, GPT) to create adaptive, personalized quizzes based on user inputs and documents. The design of the system addresses the limitations of conventional quiz platforms by offering dynamic content, instant AI-generated feedback, and performance analytics. Initial implementation and testing reveal its effectiveness: quizzes are generated swiftly and correspond with user-supplied material, while the feedback and analytics system promotes continuous learning. These features collectively aim to "assist users in learning more effectively, testing more intelligently, and growing consistently," as stated in our project mission.

Looking forward, we have several enhancements planned. Firstly, expanding the platform to mobile devices will enhance accessibility and convenience. We also plan to incorporate gamification elements (such as points, badges, and leaderboards) to further engage learners. For educators, we envision customizable quiz templates, enabling teachers to create question sets tailored to specific formats or curricula. An advanced analytics dashboard for teachers could offer real-time insights into class performance and knowledge gaps, facilitating data-driven instruction. On the AI front, future work will focus on refining the question-generation models through domain- specific fine-tuning and investigating multi- modal inputs. Lastly, we aim to conduct user studies to quantitatively assess the learning improvements enabled by the system. Through these advancements, we anticipate the platform will transform into a robust educational tool, further connecting personalized study with assessment.

## 9. References

1. ***Biancini, A., Rossi, P., & Lamberti, F.*** *(2025). Evaluating large language models for automated multiple-choice question generation in e-learning environments. International Journal of Artificial Intelligence in Education, 35(2), 211–229. https://doi.org/10.1016/j.ijai.2025.02.0 04*

2. ***Kusam, F., & Chen, M.*** *(2025). A comparative study of AI-generated and instructor-designed quizzes in higher education. Computers & Education: Artificial Intelligence, 6(1), 100123. https://doi.org/10.1016/j.caeai.2025.1 00123*

3. ***Durgungoz, A., & Durgungoz, T.*** *(2025). Gamified AI-based quizzes to enhance learning motivation and retention: A case study in online education. Education and Information Technologies, 30(4), 4551–4572. https://doi.org/10.1007/s10639-025- 12725-9*

4. ***Wang, J., & Fan, L.*** *(2025). Adaptive learning with GPT-4 integrated tutoring systems: A personalized feedback approach. IEEE Transactions on Learning Technologies, 18(3), 512–523. https://doi.org/10.1109/TLT.2025.3178 459*

5. ***Nair, S.*** *(2025). AI-driven adaptive learning systems: A review of models, applications, and challenges. Journal of Educational Computing Research, 63(1), 87–105. https://doi.org/10.1177/073563312512 3456*

6. ***Lee, C., Park, Y., & Lim, J.*** *(2025). Human–AI collaboration in question generation: Analyzing GPT-4's role in producing valid assessment items. Computers in Human Behavior: Reports, 8, 101305. https://doi.org/10.1016/j.chbr.2025.10 1305*

7. ***Chen, X., Zhang, Q., & Liu, Y.*** *(2025). Socratic Playground: GPT-4 powered intelligent tutoring for language and reasoning skills. Computers & Education, 203, 104798. https://doi.org/10.1016/j.compedu.202 5.104798*

8. ***Kulshreshtha, R., Patel, S., & Sharma, D.*** *(2025). Few-shot transformer models for personalized question generation in intelligent tutoring systems. Artificial Intelligence Review, 58(6), 3391–3415. https://doi.org/10.1007/s10462-025- 10597-2*