IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 October 2025, accepted 23 November 2025, date of publication 28 November 2025, date of current version 12 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3638427

RESEARCH ARTICLE

AI-Driven Personalized Learning and Remedial Recommendation Through Knowledge Concept-Centric Evaluation

N. PRADEESH[1], M. G. THUSHARA[1], K. ARUN KRISHNA[1], V. PRANAV[1], AND SHIVSUBRAMANI KRISHNAMOORTHY[2]

[1]Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri 690525, India
[2]Department of Computer Science and Applications, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri 690525, India

Corresponding author: N. Pradeesh (npradeesh@am.amrita.edu)

ABSTRACT This paper presents an AI-driven framework for personalized learning and remedial recommendation grounded in knowledge concept-centric evaluation. The proposed framework comprises four key modules: Knowledge Gap Identification, Adaptive Question Generation, Personalized Evaluation, and Remedial Recommendation, each designed to address individual learner needs dynamically. Retrieval-Augmented Generation (RAG) techniques are employed to create contextually aligned multiple-choice questions, while Knowledge Tracing (EKT) models continuously monitor learner progress across concepts. Experimental results based on real-world classroom deployments demonstrate significant improvements in learner competency levels following personalized interventions. Furthermore, the system supports Outcome-Based Education (OBE) principles by aligning assessments with course outcomes and learning objectives. By promoting individualized support and continuous learning, the framework advances the objectives of Sustainable Development Goal 4 (SDG 4), fostering inclusive and equitable quality education. This work highlights the potential of AI to enhance educational assessment, close learning gaps, and support scalable, competency-focused learning environments.

INDEX TERMS Personalized Learning, knowledge gap identification, adaptive question generation, deep knowledge tracing, retrieval-augmented generation, remedial recommendation, outcome-based education, AI in education, sustainable development goal 4 (SDG 4), concept-centric evaluation.

I. INTRODUCTION

In the evolving landscape of education and professional training, concept-based evaluation has become a critical mechanism for assessing learners' competencies and practical knowledge [1]. Unlike traditional grading systems that often emphasize rote memorization, knowledge concept-based evaluation prioritizes the application of knowledge, problem-solving abilities [2], and critical thinking [3]. This shift is essential for preparing individuals for real-world challenges and dynamic work environments. Nevertheless, scaling such evaluations remains difficult due to the diversity of knowledge domains and the demand for personalized evaluation strategies [4].

Traditional evaluation methods present several limitations [4]. Manual question creation is time-consuming and subject to bias, and static assessments often fail to accommodate the varied learning paces and backgrounds of learners. Scalability also becomes a critical concern when assessing large groups across diverse disciplines, limiting the timeliness and effectiveness of feedback mechanisms.

Over the past decade, artificial intelligence (AI) has emerged as a transformative force across industries, particularly within education [5]. Personalized learning systems, powered by AI, have demonstrated the ability to dynamically adjust content delivery and assessments to meet the needs of individual learners, thus enhancing engagement and mastery

The associate editor coordinating the review of this manuscript and approving it for publication was Jeffrey Jonathan Yackley.

of knowledge concepts [6]. Using Natural Language Processing (NLP) and Generative AI offers a promising solution to these challenges [7]. Through automation of knowledge concept extraction and the generation of adaptive question banks, AI can provide assessments tailored to individual learning paths [8]. This approach enables scalability while ensuring accuracy through human-in-the-loop validation processes.

The central question addressed by this study is, "How can knowledge gap identification and adaptive remedial recommendation be automated while maintaining accuracy through human feedback?" Integrating AI methodologies with advanced machine learning techniques provides a scalable, learner-centered educational framework, addressing the limitations inherent in traditional static evaluation models [9]. Such integration not only improves efficiency but also promotes fairness and adaptability, better equipping individuals for future professional demands [10].

This research adopts the principles of Outcome-Based Education (OBE) [11], [12], which emphasize the achievement of specific learning outcomes rather than focusing solely on content delivery. In the OBE framework, each course is designed around well-defined course outcomes (COs), mapped further to broader program outcomes (POs). Each CO is associated with multiple learning objectives or knowledge concepts. By integrating NLP and Generative-AI within this system, it is possible to automate the extraction of key knowledge concepts and generate adaptive assessments aligned with the desired competencies. This ensures that evaluation practices remain concept-based while directly supporting the intended educational outcomes [13], [14].

The study further explores the application of retention-augmented generation (RAG) [15], [16] to automate the generation of multiple-choice questions (MCQs), ensuring that the assessments are contextually relevant and aligned with the learning outcomes. In addition, it investigates how AI-driven knowledge gap analysis can support personalized learning pathways through adaptive assessments that identify and address individual strengths and weaknesses. Concept-based content creation techniques are also examined to generate targeted learning materials, thereby strengthening the effectiveness of knowledge-centric education.

Knowledge tracing plays a crucial role in identifying learning gaps by providing continuous, concept-level insights into learner progress [17]. In this work, an Exercise Aware Knowledge Tracing (EKT) [18] framework is utilized to dynamically model each learner's mastery across different concepts. Once knowledge gaps are identified, personalized remedial recommendations are generated using Generative AI models [19]. To enhance the reliability and relevance of these recommendations, a cross-verification process involving multiple AI modules is employed [20]. Each recommendation is aligned with the learner's current conceptual mastery level [17], providing targeted learning interventions designed to strengthen specific areas of deficiency. This personalized, adaptive process supports the broader goal of improving conceptual understanding and academic performance.

Alignment with SDG 4 – Quality Education: This research directly supports the aims of Sustainable Development Goal 4 (SDG 4) [21], which seeks to ensure inclusive, equitable, and quality education while promoting lifelong learning opportunities for all. Through AI-driven personalized assessments, automated knowledge gap detection, and remedial content generation, the framework addresses critical barriers to scalable and equitable learning. By tailoring instruction to diverse learner profiles and promoting continuous skill development, the system advances competency-focused education. Furthermore, by integrating the principles of OBE, the framework ensures alignment with real-world skills and expectations, thereby contributing to the realization of meaningful and measurable educational outcomes. In doing so, it advances both the effectiveness and fairness of educational opportunities.

A. CONTRIBUTIONS

1) AI-DRIVEN CONCEPT-BASED PERSONALIZED LEARNING

This paper proposes a modular, AI-enhanced framework centered on knowledge concepts rather than rote learning. Through the integration of Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), and Generative AI, the system automates concept extraction, multiple-choice question generation, and personalized evaluation to enable adaptive learning pathways tailored to individual needs.

2) DYNAMIC KNOWLEDGE TRACING AND REMEDIAL RECOMMENDATION

An Exercise Aware Knowledge Tracing (EKT) model is employed to continuously monitor learners' evolving knowledge states at the concept level. This enables real-time identification of learning deficiencies and facilitates the generation of personalized remedial content, with recommendations validated through a cross-verification process to ensure pedagogical relevance.

3) SCALABLE IMPLEMENTATION ALIGNED WITH OBE AND SUPPORTING SDG 4

The framework is implemented within a live academic environment through an in-house learning management system (LMS), capturing real-world learner interactions over a four-year engineering program. By aligning assessments with course outcomes and structured attainment metrics, the framework supports the principles of Outcome-Based Education (OBE) and contributes directly to the objectives of Sustainable Development Goal 4 (SDG 4) by promoting equitable, personalized learning at scale.

II. RELATED WORKS

Personalized evaluation systems, underpinned by artificial intelligence (AI), are increasingly transforming educational practices by facilitating learner-centered experiences. These systems utilize intelligent tutoring, adaptive learning platforms, and predictive analytics to discern individual learning patterns and preferences, thereby enabling the delivery of customized content and real-time feedback [22]. Machine learning techniques further enhance these capabilities by dynamically mapping learners' behavioral attributes to specific learning styles, contributing to the optimization of the overall learning process [23].

Building upon this personalized learning paradigm, recent advancements have introduced AI-driven tools for the automated generation of multiple-choice questions (MCQs) through Retrieval-Augmented Generation (RAG) frameworks. Notably, MCQGen [24] integrates large language models (LLMs) with RAG methodologies to produce MCQs that are adaptable to various educational contexts. By employing sophisticated prompting strategies, such as chain-of-thought and self-refinement techniques, these systems are capable of generating questions that are both contextually appropriate and pedagogically effective. Through the targeted addressing of common misconceptions and alignment with individual learning trajectories, such tools contribute significantly to the advancement of personalized learning [24].

Furthermore, AI-powered educational tools play a pivotal role in addressing knowledge gaps while promoting essential 21st-century skills [25], including collaboration, communication, and critical thinking. Murali et al [26] highlight how combining LLMs with structured knowledge representations enhances accuracy, explainability, and trustworthiness in medical AI. By employing predictive analytics and learner profiling, these systems support educational institutions in preparing learners to meet contemporary workforce demands. Adaptive learning platforms, in particular, tailor content and feedback according to individual learner profiles, thereby enhancing both learner engagement and educational outcomes.

A. AI IN PERSONALIZED LEARNING AND EVALUATION

Artificial intelligence has impacted individualized learning through adaptive systems that target an individual learner profile. The changing delivery of content and assessment methods, as well as the pace, contribute to better performance and learner engagement. Intelligent tutoring systems, for example, use predictive analytics to estimate learning patterns and provide specific targeted intervention for deficiencies [27]. Generative AI frameworks, including LLMs, also automate the creation of multiple-choice questions (MCQs) by making sure that the questions are relevant to the context and that the difficulty level is right [16], [24]. Techniques such as chain-of-thought prompting further align assessments with cognitive levels outlined in Bloom's Taxonomy [28].

AI-powered evaluation systems enhance learning outcomes by delivering real-time, personalized feedback. These systems map learners' behavioral attributes to learning styles, fostering the development of critical skills such as collaboration and problem-solving [29]. Adaptive platforms adjust content delivery based on learner progress, promoting inclusivity and equitable access to education [30]. Despite these advancements, challenges such as algorithmic bias and scalability persist. Future research should prioritize the integration of ethical frameworks [31] and explainable AI models to ensure transparency and fairness. Moreover, the development of multimodal content generation and robust validation techniques [32] can enhance the adaptability and inclusivity of AI-driven educational platforms [33].

While adaptive learning platforms and intelligent tutoring systems have shown promise in personalizing educational delivery, most approaches remain limited to either predictive analytics or learner-style adaptation. They rarely incorporate fine-grained concept-level knowledge gap analysis in an OBE framework, which is very important to align assessments with measurable outcomes. Our framework directly addresses this limitation by embedding knowledge tracing and OBE-based attainment analysis into the personalization process, which makes sure that it has both scalability and educational relevance.

Recent research has expanded the understanding of AI-driven adaptive learning by emphasizing both technical innovation and learner experience. Gyonyoru and J Katona [34] examined student perceptions of AI-enhanced adaptive learning systems and reported positive impacts on engagement, trust, and perceived fairness. Gyonyoru and J Katona [35] also provided a comprehensive overview of AI-based adaptive learning applications, discussing how artificial intelligence can support personalized instruction across multiple academic disciplines. In a broader context, Bond, Melissa and Khosravi [36] reviewed large-scale implementations of AI in higher education, highlighting challenges related to explainability, scalability, and ethical use. These studies collectively underscore the pedagogical significance of incorporating transparency, adaptability, and learner-centered design principles that also form the foundation of our proposed framework. Building upon these ideas in adaptive and learner-centered design, the present study integrates knowledge tracing and generative modeling to create a continuous feedback loop between assessment, remediation, and concept mastery.

B. AUTOMATED MULTIPLE-CHOICE QUESTION GENERATION AND CONCEPT MAPPING

Automated multiple choice question generation has redefined educational assessment by enabling the dynamic creation of questions aligned with individual learning paths. Question generation systems often employ concept maps, which serve as structural representations of knowledge, to facilitate comprehensive question formulation across various cognitive domains [37]. By transforming linear educational content into hierarchical structures, concept maps highlight relationships between key ideas, thereby supporting more effective knowledge assessment [2].

The integration of Retrieval-Augmented Generation (RAG) models has further enhanced multiple-choice question generation by combining retrieval-based methods with generative AI techniques [16], [38]. These approaches extract key concepts from learning materials and construct contextually appropriate questions using pre-trained language models, ensuring alignment with specific learning objectives [24], [39]. However, the generation of multiple-choice questions introduces unique challenges. Designing valid questions along with believable distractors that reflect common misconceptions remains complex [40]. Although advancements have occurred, the assurance of pedagogical suitability and curriculum alignment remains an active area of research [41].

Existing automated MCQ generation tools, such as MCQ-Gen, have advanced contextual and semantic quality, but they usually don't address pedagogical validity re-verification and Bloom's Taxonomy alignment. In addition, they do not often use cross-model verification mechanisms to ensure that the tests they make are good for teaching. In contrast, our approach integrates Retrieval-Augmented Generation with semantic deduplication and multi-model cross-verification, producing adaptive question sets that are not only contextually accurate but also educationally strong.

C. RECOMMENDER SYSTEMS FOR EDUCATIONAL INTERVENTIONS

Recommender systems play an essential role in modern educational technology by enabling personalized interventions based on individual learner profiles, knowledge states, and preferences. These systems utilize various approaches, including collaborative filtering, content-based filtering, knowledge-based methods [42], [43], tag-based methods, and hybrid models [44], to suggest learning resources and activities that support learner progress [45].

Collaborative filtering recommends resources based on patterns in learner behavior and similarities with other learners. However, these systems often face challenges such as the cold-start problem, which have been addressed through context-aware and deep learning variants [17]. Content-based recommenders focus on analyzing features of learning materials and matching them with learner profiles using semantic, attribute-based, or query-based techniques. Knowledge-based systems aggregate information about learners and content, often employing ontologies to ensure alignment with pedagogical goals [45]. Tag-based recommenders encourage learners to annotate resources, enhancing both meta-cognitive engagement and recommendation quality [46]. Hybrid models combine multiple techniques to exploit their complementary strengths, resulting in more effective recommendations [43], [47].

Recent advancements have introduced deep learning and reinforcement learning techniques into educational recommender systems [48]. Models based on long short-term memory (LSTM) networks treat recommendation as a sequential prediction problem, adapting to evolving knowledge states. Graph neural networks (GNNs) model relationships among concepts and learner mastery to support group-level recommendations. Autoencoder-based methods reconstruct sparse learner-item matrices to improve prediction accuracy [49], while relation-aware self-attention mechanisms enhance knowledge tracing and sequencing [50]. Reinforcement learning approaches, such as Advantage Actor-Critic (A2C), enable dynamic adaptation of learning paths based on real-time learner performance [51]. Other methods, including neural collaborative filtering, K-nearest neighbor, and rule-based adaptations, further refine personalization processes [45].

Despite technological advancements, challenges persist in ensuring that personalization aligns with pedagogical goals while maintaining explainability and fairness. The evaluation of educational recommender systems increasingly emphasizes not only technical accuracy but also the systems' impact on learning outcomes and learner engagement. Addressing algorithmic bias, ensuring data privacy, and developing multidimensional evaluation frameworks remain critical areas for future research [52]. As the field progresses, recommender systems are expected to play an increasingly vital role in supporting adaptive, learner-centered education that fosters motivation, engagement, and equitable opportunities for all learners.

Modern recommender systems, whether collaborative, content-based, or hybrid, improve personalization, but they mostly focus on matching learners with items instead of keeping track of concepts over time. Very few systems take into account how knowledge states change over time or how concept-aligned remedial loops work. Our framework improves this area by combining knowledge tracing with iterative remedial feedback. This makes sure that recommendations are not just static ideas but changing, concept-focused treatments that are in line with course results.

While previous research has advanced adaptive assessment, automated question generation, and recommender systems individually, they remain fragmented and rarely address integration with Outcome-Based Education (OBE). Furthermore, existing ensemble or validation methodologies do not explicitly cross-verify remedial content across heterogeneous AI models for pedagogical integrity. Our proposed approach addresses these limitations by combining EKT-based continuous knowledge tracing with AI-driven cross-verification of remedial recommendations, ensuring both conceptual alignment and instructional relevance. This integrated execution of 'RAG-based assessment', 'knowledge tracing', and 'recommender systems' extends the practical scope of AI in education, providing a reproducible model for real-world deployment.

III. METHODOLOGY

Our approach is built around a modular system that uses AI to personalize the learning experience for each learner. The process starts by analyzing course materials to extract key knowledge concepts. Based on these, the system creates customized multiple-choice questions that match the learner's current understanding and difficulty level, following Bloom's Taxonomy. As learners answer these questions, the system evaluates their performance, identifies areas where they're struggling, and offers targeted recommendations to help them improve. Each part of the system—from data collection to feedback—is designed to work together in a continuous loop, making learning more adaptive, effective, and aligned with real educational goals like those defined by Outcome-Based Education (OBE).

Figure 1 shows the proposed framework, which is a modular AI-based system designed to identify and bridge knowledge gaps in learner learning through personalized interventions. It begins with the Data Module, which collects past OBE (Outcome-Based Education) exam interactions. This data is analyzed by the Knowledge Gap Identification Module, which detects the learner's conceptual weaknesses and determines their current knowledge level. Based on this, the Recommender Module identifies the appropriate Bloom's Taxonomy Level (BTL) and target concepts for further evaluation. These inputs are sent to the Adaptive Multiple-choice Question Generation Module, which dynamically generates questions tailored to the learner's needs. The learner's responses are then evaluated through the Personalized Evaluation Module, capturing detailed insights into their understanding. These evaluation interactions are fed back into the system to continuously refine and personalize the learning experience. The entire loop ensures targeted learning support, aligning with the goals of inclusive and equitable education.

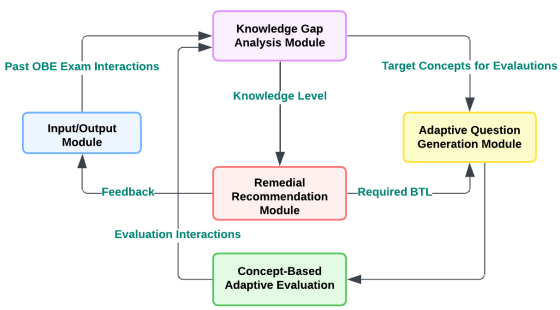

FIGURE 1. Proposed framework for personalized learning and remedial recommendation.

A. INPUT/OUTPUT MODULE

The system begins by processing a curriculum PDF document that outlines the knowledge concepts to be evaluated. This document contains course outcomes and learning objectives, which define the knowledge areas that learners must attain. Using Generative AI, the system analyzes the curriculum to extract and identify key knowledge concepts associated with the learning material [53], [54]. This automated knowledge concept extraction helps streamline the assessment process by ensuring that all critical concepts are considered for evaluation. However, since AI-generated outputs may require validation, instructors play a crucial role in reviewing the extracted knowledge concepts, making corrections, and refining them as needed. In addition, log data from e-learning platforms provides valuable insights into learner behavior, enabling systems to refine personalization strategies for enhanced engagement and retention [55]. This ensures that the knowledge concept list aligns accurately with the intended learning goals.

Once the knowledge concepts are finalized, instructors establish the evaluation criteria corresponding to each concept. These criteria define the number of questions required at different difficulty levels (Easy, Medium, Hard) and specify the pass percentage necessary for learners to progress through these levels. For instance, an instructor may require that a learner correctly answer at least 80% of the easy-level questions before advancing to the medium level.

Based on the predefined parameters, the system generates concept-based assessments by creating adaptive [56] and structured questions that correspond to the specified knowledge concept levels. This approach ensures systematic evaluation of learners' understanding while supporting personalized and scalable assessments [57]. Additionally, the use of effective concept mapping techniques aligns the assessments with prerequisite knowledge relationships, promoting a structured progression of knowledge acquisition, particularly within technical domains [58].

1) OBE-THRESHOLD AND TARGET

In our approach, learner performance within the OBE framework is evaluated using threshold and target values that are dynamically configured within the system. The Threshold in our application is typically set between 50–60%, representing the minimum level of achievement a learner must attain to meet the essential course outcomes. The Target is usually configured as approximately 20% higher than the Threshold (e.g., if the Threshold is 50%, the Target is set at around 70%), reflecting the desired level of mastery that shows strong conceptual understanding. Learners who perform below the threshold would require remedial support, while those achieving at or above the target are identified as having a proficient understanding of the subject. This systematic framework promotes a balanced emphasis on the basic competencies as well as academic excellence, thus ensuring that effective teaching strategies support the preparation of learners' readiness [59].

B. KNOWLEDGE GAP ANALYSIS MODULE

The Knowledge Gap Analysis module within our framework is designed to provide a detailed evaluation of learner performance at the concept level, facilitating the identification of specific areas where learners require targeted interventions [40]. This approach is built on the principles of Outcome-Based Education (OBE), ensuring that the evaluation process is structured, measurable, and aligned with the desired educational outcomes. In the OBE framework, each course outcome is assigned a threshold value, representing the minimum pass mark, and a target value, which defines the expected level of achievement. These parameters serve as benchmarks for determining whether learners meet, exceed, or fall short of the required competency levels for specific knowledge concepts. By comparing learners' scores against these thresholds, the system categorizes their proficiency into different levels, such as Beginner, Intermediate, or Expert, depending on their performance relative to the defined target values.

The identification of knowledge gaps is achieved by comparing learners' scores to the predefined thresholds and targets. For example, a learner scoring below the threshold value (e.g., 50%) is classified as a Beginner, while those who score between the threshold and target (e.g., 50% to 70%) are considered Intermediate, and learners exceeding the target value (e.g., above 70%) are classified as Experts. When a learner scores below the target, such as 60% in a concept with a target of 70%, the system flags this as a knowledge gap. This classification enables the provision of personalized feedback and remedial recommendations to address the identified gaps, ensuring that learners receive the necessary support to improve their understanding and achieve mastery. Additionally, the system enhances the accuracy of knowledge gap analysis by modeling the interdependencies between concepts, which allows for a more comprehensive evaluation of learners' competencies across related concepts [60].

1) KNOWLEDGE TRACING

Knowledge tracing plays a critical role in knowledge gap analysis by providing a dynamic and personalized understanding of a learner's evolving mastery across different concepts. Unlike traditional assessments that offer only static snapshots of learner performance, knowledge tracing continuously updates a learner's knowledge state based on their interactions, such as quiz responses or practice activities [61]. This enables more accurate detection of learning deficiencies and supports timely, targeted interventions.

Among the various latest knowledge tracing models, Exercise-aware Knowledge Tracing (EKT) [18] has gained attention for its ability to integrate both learner response sequences and the semantic content of exercises using deep learning techniques. Unlike traditional models or earlier RNN-based methods like DKT [62], DKVMN [63], EKT leverages textual information and contextual relationships between exercises to better understand how knowledge evolves over time. By incorporating exercise content alongside learner interaction data, EKT enables more refined tracking of concept mastery. This facilitates the creation of highly personalized remedial content and supports more targeted, adaptive learning strategies that are responsive to both what and how learners are learning.

In the Exercise-aware Knowledge Tracing (EKT) model, both the learner's activity sequence and the content of the exercises they attempt are considered to understand how their knowledge evolves over time. Each learning step involves two key elements: how the learner responded (e.g., correct or incorrect) and the actual content of the question. The EKT model combines these two inputs to track changes in the learner's understanding.

To achieve this, the model first converts the exercise text into a numerical format that captures its meaning. This is then combined with the learner's response and passed into a sequence model such as an LSTM (Long Short-Term Memory) network. The LSTM updates the learner's hidden knowledge state at each step, using both the current input and what it has learned from past steps. This allows the system to model not only how the learner performs but also what type of content they struggle with or attain.

At every time step t, the hidden state h_t summarizes the learner's knowledge based on past responses and question content. This is then used to predict how well the learner understands a set of predefined concepts. These predictions are generated through a final layer in the model that maps h_t to probabilities, each indicating how likely it is that the learner has mastered a specific concept.

For example, if the model is tracking five key concepts, the output might be:

$$y_t = \sigma(Wh_t + b) = \begin{bmatrix} 0.82 \\ 0.35 \\ 0.76 \\ 0.60 \\ 0.91 \end{bmatrix}$$

These values reflect the learner's predicted mastery level for each concept and play a key role in identifying weak areas. Concepts with mastery levels lower than the predefined OBE target value are flagged and sent to the remedial recommendation module for further support.

C. REMEDIAL RECOMMENDATION MODULE

Once knowledge gaps are identified, the system generates a personalized remedial recommendation plan to help learners address their deficiencies. For each weak concept, the system reviews the questions attempted in the most recent evaluation, along with their associated difficulty levels and Bloom's Taxonomy classifications.

Using this information, a prompt is generated and sent to a generative AI model to create targeted remedial content [64], [65]. The AI-generated recommendation includes a brief explanation of the concept, relevant learning objectives, suggested topics for review, and a set of practice problems. An example prompt used for this process is as follows:

Generate a structured remedial recommendation for the concept "Algorithmic Problem Solving

> *and Array Manipulation" for a learner who demonstrates 45% concept mastery. The response should include five sections: Learning Objectives, Recommended Topics to Revise, Remedial Explanation, Two Practice Activities and Concept Gap Rationale, which will describe the specific weaknesses detected by the knowledge-tracing model and the reasoning behind the generated remedial content.*

To ensure the accuracy and pedagogical relevance of the generated content, the system employs a cross-verification mechanism. The initial output is reviewed using additional generative AI models—such as Gemini or DeepSeek [20]—to validate whether the explanation aligns with the intended Bloom's level and concept difficulty. These models evaluate the structure, clarity, and instructional value of the content. More details about the AI data verification is explained in the sections III-D2 and III-D3.

Each recommendation is structured into five essential parts: (1) **Learning Objectives** that define the specific knowledge concepts to be acquired, (2) **Review Topics** that guide learners on what foundational concepts need to be revisited, (3) **Remedial Explanation**, 4) **Practice Activities** that reinforce learning through hands-on exercises, and (5) **Concept Gap Rationale** that explains to the learner which gaps were detected by the knowledge-tracing model and why the particular recommendation was generated.

For instance, Figure 2 shows a sample recommendation; when a learner fails to answer a question related to *Algorithmic Problem Solving and Array Manipulation*, a detailed remedial recommendation is generated automatically. The example recommendation includes a breakdown of necessary array operations, sorting strategies, and problem-solving techniques needed to solve similar problems in the future. Each remedial suggestion is comes with a brief explanation outlining the underlying concept gap detected by the knowledge-tracing model to support transparency and learner understanding.

These recommendations are generated using OpenAI's language models through API integration, ensuring natural language clarity and contextual accuracy [64]. To ensure reliability and correctness, all outputs are cross-verified using additional AI models and manual review. This hybrid validation strategy ensures the feedback is both pedagogically sound and technically accurate, enabling a more effective learning process for the learners.

After receiving the remedial recommendation, the learner studies the suggested content and attempts the included practice problems. Once they feel ready, they complete a self-assessment that consists of a new set of questions covering the same concepts but with different structures. This ensures that any improvement in performance reflects genuine understanding rather than memorization. The results of this self-assessment are then used to update the learner's knowledge state through the EKT model. If the learner still

Remedial Recommendation for Concept: Algorithmic Problem Solving and Array Manipulation

1. Learning Objectives:

- Understand basic array operations (traversal, indexing, swapping).
- Learn how to apply simple sorting or selection techniques (e.g., finding top N elements).
- Practice maintaining the relative order of elements after a targeted modification.

2. Recommended Topics to Revise:

1) **Array Traversal and Indexing** leftmargin=1.5em,nosep
   - Accessing and modifying elements using loops.
   - Understanding 0-based indexing in arrays/lists.
2) **Sorting & Selection Algorithms**
   - Basics of sorting (e.g., bubble sort or using built-in sorting functions).
   - Finding the top 3 elements using sorting or comparison.
3) **Reconstruction of Arrays**
   - Creating a new array with specific elements at desired positions.
   - Using slicing or loop-based reconstruction.
4) **Problem-Solving Patterns**
   - Decomposing a problem into steps.
   - Writing pseudocode before implementation.

3. Remedial Explanation:

- Think of an algorithm as a set of steps to solve a problem—just like following a recipe. In programming, we often write algorithms to work on arrays, which are like ordered shelves that hold values (like numbers or names). For example, to find the largest number in a list (array), the steps could be:
  - Assume the first number is the largest.
  - Compare it with every other number.
  - If a bigger number is found, update your assumption.
  - After checking all, the largest is the one left.
- This is called a linear search algorithm. You use the same idea to search, insert, delete, or update values inside an array.

4. Practice Activities:

- **Problem 1:** Write a program to find the largest number in an array.
  Input: `[10, 25, 5, 60, 12]` *Expected Output:* `60`
- **Problem 2:** Insert a number at a specific position in an array.
  Input: `Array = [3, 8, 9, 15]`, Insert: `6`, Position: `2`
  Expected Output: `[3, 8, 6, 9, 15]`

5. Concept Gap Rationale: This recommendation was generated because the knowledge-tracing model detected repeated errors in questions involving array traversal, indexing, and basic element manipulation. These gaps suggest the learner needs focused practice on foundational array operations before advancing to more complex problem-solving tasks.

FIGURE 2. Remedial recommendation for concept: Algorithmic problem solving and array manipulation.

hasn't reached the target mastery level, the system generates a new round of refined recommendations—possibly including more detailed explanations, simplified examples, or additional practice activities. This creates an ongoing learning loop where feedback from each round informs the next, helping learners gradually build up their understanding until they reach the required level of competency.

D. ADAPTIVE QUESTION GENERATION (RAG MODEL)

The adaptive question generation module leverages a Retrieval-Augmented Generation (RAG) approach to produce curriculum-aligned multiple-choice questions that vary in difficulty and remain contextually relevant to instructor-defined learning objectives. Curriculum documents are processed to extract key concepts and learning outcomes, which are converted into semantic embeddings and stored in a vector database to enable fast, context-aware retrieval during question generation. This design ensures that questions remain strictly within the scope of the provided materials, maintain balanced difficulty across

Bloom's Taxonomy levels, and provide sufficient variability for fair and adaptive assessments. Detailed descriptions of preprocessing steps, vectorization workflow, conversational retrieval chain, and persistent memory structures are provided in Appendix A.

1) SIMILARITY-BASED DEDUPLICATION OF MCQS

To maintain and ensure the quality of the question bank with regard to non-redundancy, each of the newly generated questions is compared against the existing questions stored in the global memory. The comparison is made through semantic similarity, where the embeddings of the new questions are compared to those stored. Cosine similarity is employed as the measure for comparison, and all new questions that score above a threshold (for example, 0.85) are labeled as duplicates and discarded. Only unique questions are put back into the global memory for future comparisons. Thus, high-quality, non-repetitive questions will be populated such that they can still satisfy the intended learning objectives. This last filtered set of questions may then be uploaded to a web interface for manual review and retrieval or returned as a JSON-formatted response for future use. The validation step is important in ensuring the integrity of the question bank, which ensures that the questions generated are unique and aligned with the educational goals of the system.

The final set of filtered questions is either presented on a web interface for manual review or returned as JSON-formatted responses for further use. This validation step plays a critical role in maintaining the integrity of the question bank, ensuring that the questions generated are both distinct and aligned with the educational goals of the system.

2) AI-BASED QUESTION VERIFICATION AND REFINEMENT

For multiple-choice question generation, OpenAI APIs are used to ensure layered diversity and representative questions at each knowledge concept level. The AI-generated questions will now specify knowledge concepts and evaluation requirements set by the instructors. The system then verifies that the questions produced by AI are high quality and accurate through an additional verification process. The same OpenAI model is used in prompt correction to verify that generated questions are well-formed, grammatically correct, and contextually relevant. This iterative process improves the quality of relevant questions for further validation.

3) CROSS VERIFICATION

To guarantee the relevance and complexity of the questions produced, we use a cross-verification method that involves the use of several Generative AI models. The system-generated questions are then subjected to secondary testing using Gemini AI and DeepSeek AI, which determine if the questions meet the required knowledge concept level and the set complexity. These models assess the questions against the specified learning objectives so that they are suitably challenging—neither too simple nor too complex—depending on the level of targeted knowledge concept.

Algorithm 1 Personalized Learning and Recommendation

Input: Course Curriculum (PDF), OBE Exam Result
Output: Remedial Recommendations and Personalized Evaluation
Step 1: Data Preparation
1.1 Convert curriculum PDF to text.
1.2 Apply RAG to curriculum and store in vector database
Step 2: Knowledge Gap Identification
2.1 Knowledge Tracing: Generate knowledge state and identify knowledge gap.
Step 3: Remedial Recommendation
3.1 Generate recommendations for the identified concepts using AI
3.2 Cross-verify using other AI models and deliver to learner
Step 4: AI-Driven Multiple-choice Question Generation
4.1 Generate concept-based questions using RAG
4.2 Check for duplicate questions by comparing the semantic similarity.
4.3 Cross verify the questions, validate BTL and relevance
Step 5: Personalized Evaluation
5.1 Select questions based on knowledge gap.
5.2 Generate exams using instructor-defined rules
5.3 Deliver exam to the learner
5.4 Pass exam result to Step 2 to identify knowledge gap.
5.5 Continue learning cycle

By incorporating multiple AI models into the verification process, we enhance the fairness, accuracy, and adaptability of the assessment system. This approach ensures that learners are evaluated using high-quality questions that are well-balanced and appropriate for their level of understanding, ultimately contributing to more accurate and reliable assessments.

E. CONCEPT-BASED ADAPTIVE EVALUATION

The Concept-Based Adaptive Evaluation system guarantees a knowledge concept-based evaluation to every learner by dynamically creating assessment questions that are specific to his or her knowledge concept level. Unlike static evaluations, this system provides every learner with a specific set of questions for every knowledge concept level, thus enhancing the adaptability and equity of the evaluation process. By structuring the evaluation into different levels, learners progress through adaptive stages based on their mastery of specific concepts, providing a personalized learning experience that is closely aligned with their capabilities [66]. A higher difficulty level is introduced only once the learner successfully clears the current level, ensuring a structured and progressive approach to assessment.

The success or failure at every level of knowledge concept is determined based on instructor-established assessment criteria, including the number of questions needed at every level of difficulty (Easy, Medium, Hard) and the percentage required to pass every level. Learners are grouped into Beginner, Intermediate, and Expert categories based on the highest level they pass. Candidates who don't succeed at the Easy level are categorized as Not Competent, while candidates who pass at the Easy level but not the Medium level are placed as Beginners in that given knowledge concept. This structured evaluation process ensures accurate assessment of each learner's proficiency and provides a clear learning path for further improvement in the respective knowledge concepts.

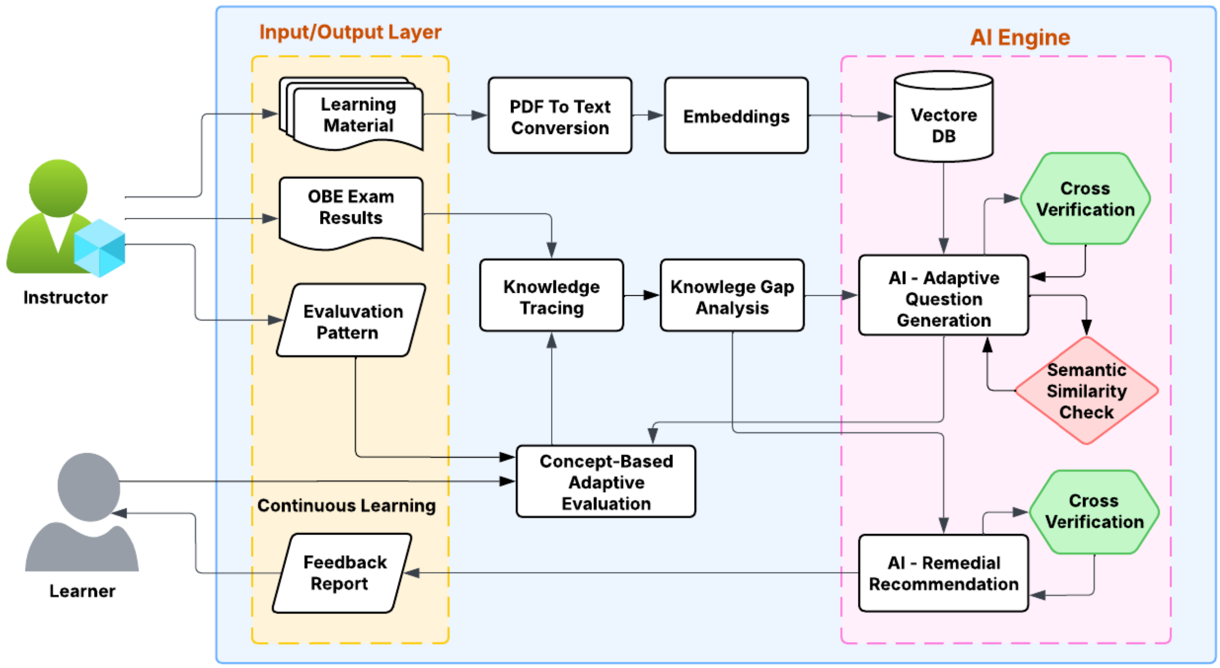

FIGURE 3. System architecture of the proposed AI-driven personalized learning framework integrating Retrieval-Augmented Generation (RAG), Exercise-Aware Knowledge Tracing (EKT), and generative AI-based remedial recommendation modules. The proposed feedback loop demonstrates measurable impact, achieving a high learner satisfaction rate (Figure 9) and a remarkable improvement in concept-wise competency levels (Figure 7). This architecture ensures both pedagogical alignment and adaptive personalization through transparent and explainable AI mechanisms.

F. FEEDBACK AND CONTINUOUS LEARNING

Learners can engage in an iterative and continuous learning process until they successfully attain the targeted concepts and its levels. Initially, they must provide their OBE-based evaluation scores, which serve as the foundation for the system to generate personalized recommendations using the exam data. Once these recommendations are studied, learners can perform a self-evaluation to test their progress. Based on the results of this self-assessment, the system will generate new learning recommendations and curate a fresh set of evaluation questions. These new questions are crafted by considering the learner's previous performance, and they assess the same concepts through varied question structures to ensure the learner's success is based on genuine understanding rather than guesswork. Deliberate practice activities focusing on weak areas improve learning outcomes by encouraging learners to systematically address deficiencies rather than revisiting mastered content unnecessarily [41]. This cycle of learning, self-evaluation, and feedback-driven question generation continues until the learner feels confident and satisfied with their mastery of the concepts.

G. OVERVIEW OF THE PROPOSED METHODOLOGY

Figure 3 shows the system architecture of the proposed AI-driven personalized learning framework integrating Retrieval-Augmented Generation (RAG), Exercise-Aware Knowledge Tracing (EKT), and Generative AI–based Remedial Recommendation modules. The purpose of this architecture is not only to automate question generation but also to ensure that assessments remain pedagogically aligned and adaptive at the concept level. This distinguishes our approach from conventional MCQ pipelines that focus on automation without addressing instructional validity and interpretability.

The proposed algorithm (Algorithm 1) outlines an end-to-end pipeline that integrates curriculum processing, concept-level knowledge tracing, AI-driven remedial recommendation, and adaptive assessment into a continuous learning loop. It starts by converting curriculum content into searchable embeddings, then maps learner performance from OBE evaluations to corresponding concepts to identify weak areas using an Exercise-Aware Knowledge Tracing (EKT) model. Personalized remedial content is generated and cross-verified by multiple AI models, followed by concept-aligned MCQ generation to support adaptive evaluations. Learner results from these evaluations feed back into the knowledge-tracing model to iteratively refine recommendations and promote mastery. A full step-by-step workflow and detailed algorithmic description are provided in Appendix B.

IV. EXPERIMENTAL SETUP AND EVALUATION

The experiments are conducted within the context of a four-year Computer Science And Engineering program that is designed based on the Outcome-Based Education (OBE) framework. The OBE exam data was captured using an in-house LMS application called AMPLE [67], which is specifically designed to support OBE. AMPLE facilitates structured mapping of questions to course outcomes and learning objectives, while also enabling detailed performance analysis based on OBE metrics. The entire analysis presented in this study was conducted on a live, running batch of engineering learners, ensuring that the evaluation reflects real-world academic conditions and learner behaviors.

A. INPUT DATA

1) COURSE CURRICULUM

The primary dataset consists of the complete curriculum document covering 70+ courses across the four-year program. This document includes detailed information such as course outcomes, associated learning objectives, and the key concepts taught in each course. The curriculum is initially provided in PDF format, which is then converted to text and processed using a Retrieval-Augmented Generation (RAG) pipeline. Through this process, a vector database is created from the textual content, enabling efficient concept retrieval and supporting the generation of concept-aligned evaluation questions.

2) KNOWLEDGE TRACING DATASET

For training the knowledge tracing (KT) model, we utilized Outcome-Based Education (OBE) exam data collected from a complete 4-year Computer Science and Engineering (CSE) undergraduate program. The learners whose data were used have already completed the program, and hence the dataset provides a comprehensive historical view of their learning journey. As shown in Table 1, the dataset contains 382 unique concepts and performance data from 1,276 learners, resulting in a total of 359,681 recorded learner-concept interactions. On average, each learner attempted around 281 questions, covering approximately 64 concepts. The average attainment percentage stands at 61.67%, indicating that each learner, on average, answered 61.67% of the attempted questions correctly. This well-rounded dataset was used to train the KT model, which is now being leveraged to estimate the knowledge state of current live learners in real-time.

TABLE 1. OBE exam dataset for knowledge tracing model (EKT) training.

Item	Count
Concepts	382
Learners	1276
# Records	359681
Avg. questions attempted per learner	281
Avg. concepts attempted per learner	64.06
Avg. attainment percentage for learner	61.67%

3) RECOMMENDATION MODULE ANALYSIS DATASET

For analysing the Recommendation module performance, we focused on a currently running Computer Science Engineering course titled "Procedural Programming using C." The dataset comprises exam performance data from a live engineering batch of learners, including their Internal and External exams. Each record in the dataset contains the exam questions, the associated course outcomes, the specific concepts covered, and the marks scored by each learner for each question. We also identify the threshold and target values defined for each course outcome based on the OBE framework. This allows us to determine the required proficiency level for each question and assess whether individual learners have successfully attained the target outcome levels. The dataset thus serves as a basis for knowledge-tracing and knowledge gap identification for each learner.

The Table 2 gives the distribution of the dataset. The dataset includes information from 262 learners, covering their responses to 16 unique exam questions that are mapped to 5 key concepts. Each question is tagged with its associated course outcome and concept. In total, there are 1032 records, representing individual learner responses to different questions. On average, each learner attempted 9 questions, spanning approximately 4.7 concepts. The average attainment percentage across learners stands at 65%, indicating that, on average, learners were able to successfully attain the targeted knowledge concepts in 65% of the questions they attempted.

This dataset enables fine-grained analysis of concept mastery and individual learner progress, serving as the foundation for knowledge tracing, identifying learning gaps, and evaluating the effectiveness of outcome-based teaching strategies in a real classroom setting.

TABLE 2. OBE exam dataset for recommendation analysis.

Item	Count
Concepts	5
Learners	262
Unique Questions	16
# Records	1032
Avg. questions attempted per learner	9
Avg. concepts attempted per learner	4.7
Avg. attainment percentage for learner	65%

4) EVALUATION PATTERN

Along with OBE Exam result, instructor also provides the rules for knowledge concept level attainment by specifying the minimum number of questions required to evaluate a knowledge concept and the pass percentage necessary to advance. These configurations can vary based on the course's importance. For instance, in optional courses, the instructor might set the configuration to "1 and 100," meaning that a single question is sufficient for evaluation, and the learner must answer it correctly to progress to the next level. In contrast, for core courses, the instructor may require a configuration such as "5 and 80," indicating that five questions will be administered per knowledge concept level, and the learner must answer at least four correctly (80%) to move forward. If a learner fails to meet the criteria, they must retake the assessment following the remedial recommendations provided by the system.

B. METRICS FOR EVALUATION

To evaluate the effectiveness and reliability of the proposed AI-driven knowledge concept evaluation system, we employed a combination of qualitative and quantitative metrics. The analysis was performed across three key areas: question generation accuracy, knowledge gap identification, and remedial learning effectiveness.

1) MULTIPLE-CHOICE QUESTION GENERATION ACCURACY

To evaluate the quality and effectiveness of the generated questions, we employed a comprehensive set of metrics focusing on semantic accuracy, readability, diversity, and relevance. Questions were generated using multiple Generative AI models, including OpenAI, Gemini AI, and DeepSeek. These questions were further refined and cross-verified based on their alignment with pedagogical standards—specifically considering their relevance, complexity, and Bloom's Taxonomy Level (BTL). A comparative analysis was also conducted against questions authored by real teachers to assess the pedagogical appropriateness and practical applicability of the AI-generated content.

2) KNOWLEDGE GAP IDENTIFICATION ANALYSIS

A knowledge gap matrix was constructed, which lists the percentage of learners who attained or failed to attain each identified knowledge concept. This analysis provided a clear overview of the distribution of competencies among learners and helped identify high-risk areas where intervention was necessary. The matrix served as a validation for the system's ability to correctly classify learner knowledge concept levels based on evaluation results.

3) REMEDIAL EFFECTIVENESS EVALUATION

To assess the impact of the remedial recommendations, we compared learner scores before and after undergoing AI-driven remedial learning. A comparative table was created showing improvements in concept-level performance. This evaluation helped determine the effectiveness of personalized learning material in addressing the learner's deficiencies and improving their understanding of difficult concepts.

V. RESULTS AND DISCUSSION

Based on our analysis, the results are organized into three main categories: *MCQ Generation, Knowledge Tracing,* and the *Recommendation Module.* The detailed outcomes for each of these components are discussed below.

A. EVALUATION OF MCQ GENERATION USING RAG

This section presents two sets of results. First, we evaluated various AI models to identify the most suitable one for multiple-choice question generation, while others were used for verification purposes. The second set of results compares our proposed model with other popular question generation tools. A detailed explanation of both sets is provided below.

1) GENERATIVE AI MODELS COMPARISON

Table 3 presents a comparative evaluation of multiple-choice question generation performance across three generative AI models—OpenAI (O), Gemini (G), and DeepSeek (D)—over five programming-related concepts from another engineering course *Principles of programming languages*. The concepts we used for the analysis are *Encapsulation, Class, Data Hiding, Access Modifiers, and Getters and Setters.* Our objective was to generate teacher-style questions, so we used a controlled scenario provided by a human teacher and asked each model to generate questions for the same. To assess the quality and alignment of these generated questions with human-generated ones, we performed semantic similarity analysis using pre-trained language models: XLNet [68], Word2Vec [69], S-BERT [70], and Gensim [71]. These models converted both AI-generated and human-generated questions into embeddings, and we calculated cosine similarity scores between them. Higher similarity indicates closer alignment with human intent and phrasing.

From the results, it's evident that the OpenAI-generated questions achieved the highest semantic similarity scores for the majority of concepts. For instance, OpenAI scored as high as 0.965 for "Getters and Setters" and 0.978 for "Encapsulation" using XLNET, demonstrating a strong alignment with human-generated questions. While Gemini and DeepSeek also produced reasonably good results, their scores were comparatively lower. Based on these findings, we selected OpenAI as our primary engine for multiple-choice question generation, with Gemini and DeepSeek serving as secondary models for cross-verification. This multi-model validation approach helps ensure pedagogical accuracy and consistency in AI-generated evaluation content.

2) COMPARATIVE EVALUATION OF QUESTION GENERATION MODELS

The proposed *RAG-Cross Verified Multiple-Choice Question Generation* model was benchmarked against two popular question generation systems: *MCQGen* [72] and *Kristiyan Vachev QG* [73]. The evaluation metrics used in this study include:

a: AVERAGE SEMANTIC SIMILARITY (S-BERT)

Measures how closely the AI-generated question aligns in meaning with reference questions. We used S-BERT for semantic similarity evaluation because it provides state-of-the-art sentence-level embeddings that capture the meaning of text beyond simple keyword matching. Unlike traditional methods such as XLNET or Word2Vec, S-BERT is fine-tuned on natural language inference (NLI) and semantic textual similarity (STS) tasks [70], making it highly effective for comparing the semantic closeness of two sentences or questions.

TABLE 3. Semantic similarity comparison of AI-generated questions using multiple generative AI models (O: OpenAI, G: Gemini, D: DeepSeek) against teacher-crafted questions across five key concepts using multiple embedding models (XLNet, Word2Vec, S-BERT, and Gensim).

Concept	XLNET			Word2Vec			S-BERT			Gensim		
	O	**G**	**D**	**O**	**G**	**D**	**O**	**G**	**D**	**O**	**G**	**D**
Encapsulation	**0.978**	0.976	0.972	**0.8490**	0.7464	0.7014	**0.703**	0.414	0.385	**0.1157**	0.1104	0.0830
Class	0.972	**0.975**	0.966	**0.7998**	0.7400	0.6995	**0.444**	0.305	0.308	**0.0975**	0.0363	0.0287
Data Hiding	**0.966**	0.964	0.955	0.8159	**0.8161**	0.7798	**0.586**	0.486	0.412	**0.0743**	0.0171	0.0281
Access Modifiers	0.972	**0.976**	**0.976**	**0.8680**	0.8336	0.7979	0.464	**0.514**	0.482	**0.0868**	0.0439	0.0552
Getters and Setters	**0.965**	0.962	0.955	0.8352	**0.8947**	0.7448	0.621	**0.649**	0.395	0.1151	**0.2142**	0.1110

b: READABILITY SCORE (FLESCH-KINCAID)

Assesses how easy it is to read and understand the generated questions. The Flesch-Kincaid Readability Score [74] is a widely used metric to assess how easy a piece of text is to read and comprehend. It calculates a score based on the average number of syllables per word and the average number of words per sentence. Higher scores indicate simpler, more accessible language, which is crucial in educational contexts where clarity directly affects learner engagement and understanding.

c: DIVERSITY SCORE

To evaluate the lexical and structural diversity of AI-generated questions, we employed the Self-BLEU metric, which measures how similar each generated question is to the others in the same set. Specifically, we used the BLEU implementation from SacreBLEU [75] and calculated the BLEU score of each question by treating it as a candidate and the remaining questions as references. The average of these individual BLEU scores provides a Self-BLEU score, where lower values indicate higher diversity, as the generated questions are less similar to one another. To make the interpretation easier, we normalized the Self-BLEU by dividing it by 100, resulting in a Diversity Score, where smaller values suggest less redundancy and better variation across the generated content. This metric is particularly important in multiple-choice question generation tasks to ensure that the model is not repeating similar structures or phrases, thereby promoting richer learning experiences.

d: RELEVANCE (1–5 SCALE)

To assess the contextual relevance of AI-generated questions, we compared them with a set of teacher-authored questions using sentence-level semantic similarity. Specifically, we used the Sentence-BERT model *(all-MiniLM-L6-v2)*, which is designed for fast and accurate sentence embedding and similarity computation [70]. Each generated question and its corresponding teacher question were encoded into dense vector representations, and the cosine similarity between these vectors was calculated using *pytorch_cos_sim*. These similarity scores, which range from –1 to 1, were then mapped to a 5-point relevance scale through a heuristic: scores above 0.9 were rated as 5 (highly relevant), while those below 0.6 were rated as 1 (least relevant). This mapping helps interpret raw similarity scores in an intuitive and educationally meaningful way. The average relevance score across all question pairs provides a quantifiable measure of how well the AI-generated content aligns with expert-designed assessments in terms of learning objectives and conceptual fidelity.

The performance comparison presented in Table 4 clearly demonstrates the effectiveness of the proposed RAG-Cross Verified Multiple-Choice Question Generation model over other baseline models. It achieves the highest semantic similarity score (0.5642), indicating that the generated questions closely align in meaning with those created by human educators. Additionally, it records the best readability score (10.53), showing that the questions are easier for learners to understand—an essential factor in learner engagement and comprehension. The model also achieves the lowest diversity score (0.12), reflecting consistency and reduced redundancy in question phrasing. While the relevance score (1.56) is competitive with other models, it still maintains strong alignment with the learning context. These quantitative results collectively validate the model's ability to generate clear, meaningful, and educationally relevant questions, supporting its suitability for adaptive learning environments where question quality is critical.

B. EFFECTIVENESS OF KNOWLEDGE TRACING

In our proposed solution we have used the Exercise Aware Knowledge Tracing (EKT) [18], which helps to monitor and predict a learner's performance by updating their knowledge state dynamically at each step in the learning process. The model takes as input a sequence of learner interactions—each indicating a question (or concept) and whether it was answered correctly—and predicts the probability of the learner correctly answering future questions. We used concept attainment data to train the model, splitting 75% of the data for training and 25% for testing. The training was conducted over 12 epochs to observe the model's ability to learn and improve continuously over time.

The EKT architecture extends traditional knowledge tracing by incorporating both learner interactions and the semantic content of exercises. The model includes an embedding layer that encodes each exercise's textual features and learner response (correct/incorrect). The embedding dimension is set to 100. Deep learning KT models such as EKT [18]

TABLE 4. Comparative evaluation of MCQ generation models (RAG-Cross Verified QGen, MCQGen, and Kristiyan Vachev QG) Using four metrics: Semantic similarity (S-BERT), Readability (Flesch-Kincaid), Diversity (Self-BLEU), and relevance score.

Model	Avg. Semantic Similarity	Readability Score (F-K)	Diversity Score (↓ better)	Relevance (1-5)
Proposed RAG-Cross Verified QGen	**0.5642**	**10.53**	**0.12**	**1.56**
MCQGen	0.3893	8.59	0.39	1.36
Kristiyan Vachev QT	0.4349	9.78	0.38	1.24

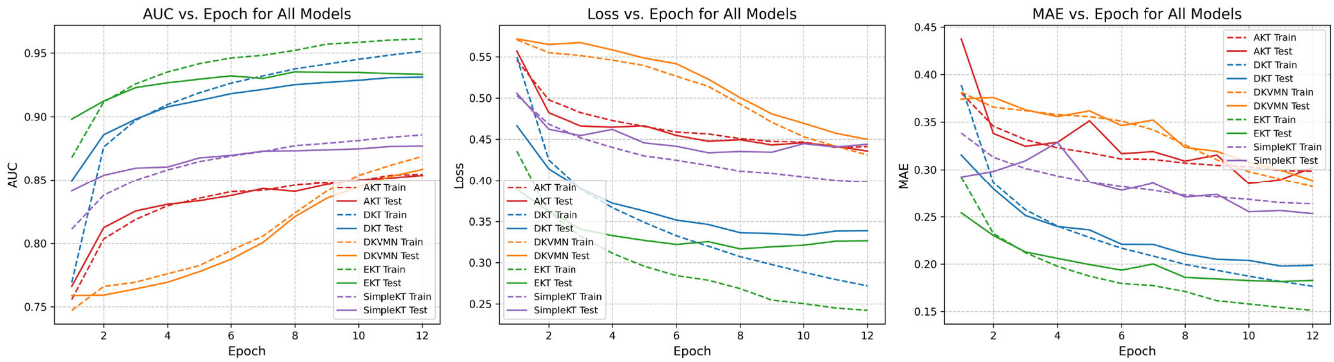

FIGURE 4. AUC progression over training epochs for DKT, DKVMN, EKT, AKT and SimplKT models. The graph demonstrates that EKT achieves consistently higher AUC values compared to other models, indicating its ability to model and predict learner knowledge states accurately.

TABLE 5. Performance comparison of knowledge tracing models (EKT, DKT, DKVMN, AKT, SimpleKT) Using LOSS, AUC, and MAE metrics on training and testing datasets.

Model	Train			Test		
	LOSS	AUC	MAE	LOSS	AUC	MAE
AKT	0.4412	0.8546	0.2984	0.4358	0.8537	0.30122
DKT	0.2718	0.9514	0.1766	0.3387	0.9311	0.1987
DKVMN	0.4312	0.8688	0.2820	0.4500	0.8585	0.2882
EKT	**0.2421**	**0.9612**	**0.1512**	**0.3266**	**0.9334**	**0.1828**
SimpleKT	0.3982	0.8857	0.2637	0.4441	0.8770	0.2532

and DKVMN [63] experimented with multiple dimension values for the vector and chose 100 as the most effective size, allowing the model to learn a meaningful representation of each concept interaction (concept + correctness). These embeddings are then passed through an LSTM layer with 100 hidden units to capture the temporal sequence of learning behavior. A dense output layer with sigmoid activation produces probabilities indicating the learner's mastery of each concept.

A dropout rate of 0.3 is applied to prevent overfitting, and the model is trained using the Adam optimizer with a learning rate of 0.001. The training is done in batches of size 32, and the performance is evaluated using AUC (Area Under the ROC Curve). The model uses binary cross-entropy as the loss function, which is suitable for binary classification problems like predicting whether a learner will answer a question correctly. This setup enables the EKT model to effectively trace the evolving knowledge state of a learner and provide valuable insights into their learning progress.

Figure 4 illustrates the AUC performance of the popular knowledge tracing models DKT [62], DKVMN [63], EKT [18] AKT [61] and SimpleKT [76] models over 12 epochs for both training and test datasets. However, after the 12th epoch, we have noticed that the models are overfitting on the given dataset. As shown in the figure4, EKT and DKT consistently achieve higher AUC values compared to other knowledge tracing models throughout the training process in the given OBE dataset. Since EKT is slightly better performance than DKT, we have chosen EKT for further analysis.

The Figure 5 shows the heatmap, visualizing the knowledge state progression of a single learner over time as they attempt different questions. The x-axis represents each question attempted, with the corresponding concept and whether the learner got it correct (1) or incorrect (0). The y-axis lists all the concepts being tracked: —*Fundamentals and System Architecture; Operators, Pointers and Control Structures; Algorithmic Problem Solving and Array Manipulation; String Manipulation and Text Processing; Data Structures, Memory Management, and File Handling*

As the learner answers each question, the EKT model updates their knowledge state, predicting the likelihood of them answering future questions correctly for each concept. These values (shown in the heatmap cells) range from 0 to 1, where a higher value indicates stronger mastery or confidence in that concept. The final column of the heatmap (i.e., the last step) summarizes the learner's current knowledge state for each concept, which is then used by the model to predict their upcoming performance on future questions. This ability to dynamically trace knowledge across time is the core advantage of the EKT model in personalized learning systems.

The Figure 6 illustrates the learner's concept-wise mastery level as predicted by the trained EKT model. This visualization represents the learner's last known knowledge state, indicating their current level of understanding across various knowledge concepts. For example, the learner

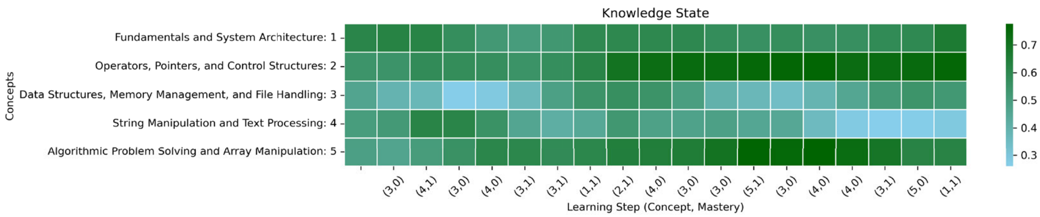

FIGURE 5. Knowledge state evolution of a learner across concepts using EKT – The heatmap shows how the learner's mastery levels for each concept update over time as they attempt questions. Each column corresponds to an interaction, and the final column reflects the predicted current knowledge state used for future performance prediction.

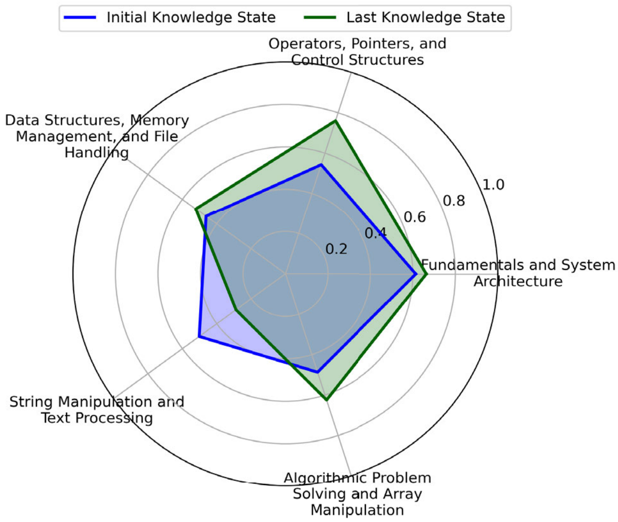

FIGURE 6. Learner's concept-wise mastery levels generated from EKT model representing the initial and final knowledge state.

demonstrates strong mastery in topics like*Fundamentals and System Architecture;* and *Operators, Pointers and Control Structures;* while showing weaker understanding in *String Manipulation and Text Processing;*. Such insights are crucial for personalized learning, as they help identify areas where the learner requires further support. Recommendations for remedial learning, additional exercises, or revision content are provided based on these mastery scores to enhance the learner's progression effectively.

C. EFFECTIVENESS OF RECOMMENDATION MODULE

Following the integration of the remedial recommendation module, a comparative analysis of learner competency levels across five core computer science domains revealed a consistent and positive shift. The stacked bar chart (Figure 7) illustrates a notable reduction in the proportions of learners classified as Not Competent and Beginner, alongside a corresponding increase in those achieving Intermediate and Expert levels. This trend was observed uniformly across all five concepts—*Fundamentals and System Architecture; Operators, Pointers and Control Structures; Data Structures, Memory Management, and File Handling; String Manipulation and Text Processing; and Algorithmic Problem Solving and Array Manipulation;*—indicating a broad and inclusive impact of the intervention.

The results suggest that the module effectively addressed learning gaps and contributed to overall skill development in foundational areas of computer science. The increase in higher competency levels across all topics reflects an enhanced understanding of the subject matter among learners. This overall improvement underscores the potential of targeted remedial interventions to elevate learner performance, promote deeper conceptual clarity, and support improved educational outcomes at scale.

The radar chart (Figure 8) visualization presents a comprehensive comparison of learners' knowledge levels across five core concepts—'Algorithmic Problem Solving and Array Manipulation'; 'Data Structures, Memory Management, and File Handling'; 'Fundamentals and System Architecture'; 'Operators, Pointers and Control Structures'; and 'String Manipulation and Text Processing'—before and after the application of personalized learning recommendations. Each chart illustrates the knowledge progression of 30 learners, with individual performance plotted in terms of percentage attainment. The observed trend across all five concepts reveals a notable shift from lower to higher performance bands, indicating an overall improvement in conceptual understanding. This improvement suggests that the recommendation strategy effectively supported personalized learning, helping learners enhance their mastery in specific areas. The visual differentiation between the pre- and post-recommendation plots provides an intuitive understanding of the intervention's impact, highlighting the potential of targeted feedback in academic performance enhancement.

We conducted a paired-sample t-test to determine the statistical significance of the difference between the pre-test and post-test scores for the same learners. It helps us to make sure that the changes we see are not just random and are really due to the intervention of the recommendation module. Table 6 summarizes the paired-sample t-test results for the five concepts assessed before and after the recommendation module. All concepts show statistically significant improvements ($p < 0.0001$), indicating that post-test scores were consistently higher than pre-test scores across the board. The largest mean gain was observed in FSA (+22.47 points), followed closely by APS (+18.50) and DS (+15.97), suggesting that the intervention had particularly strong effects

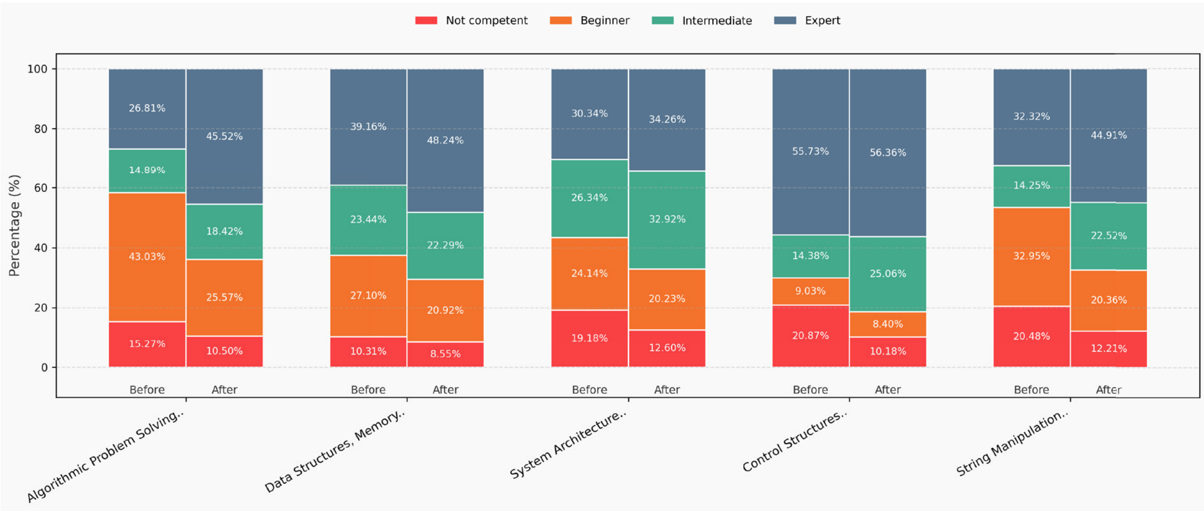

FIGURE 7. Comparison of learner competency levels before and after remedial recommendation across five concept domains. Each domain displays a stacked breakdown of learners classified into four levels—Not Competent, Beginner, Intermediate, and Expert—highlighting the overall improvement in proficiency following the recommendation module.

FIGURE 8. Radar chart showing concept-level mastery improvements for a batch of 30 Learners before and after applying AI-generated remedial recommendations across five core concepts.

TABLE 6. Paired-sample *t*-test results comparing pre-test and post-test scores for each concept.

Concept	t-statistic	p-value	Mean Pre-test	Mean Post-test	Mean Diff
Fundamentals and System Architecture	-8.86	< 0.0001	45.10	67.57	22.47
Operators, Pointers and Control Structures	-8.61	< 0.0001	45.86	60.14	14.28
String Manipulation and Text Processing	-5.70	< 0.0001	40.25	51.45	11.21
Data Structures, Memory Management, and File Handling	-6.10	< 0.0001	52.88	68.85	15.97
Algorithmic Problem Solving and Array Manipulation	-6.75	< 0.0001	39.65	58.15	18.50

on these concepts. The substantial negative t-values reflect that post-test scores exceeded pre-test scores (calculated as Pre – Post), reinforcing the effectiveness of the intervention in improving learner performance.

TABLE 7. Survey questions for evaluating AI-generated personalized recommendations.

Section A: Usefulness & Relevance-focused on whether the remedial recommendations accurately addressed the learners' learning gaps and recent mistakes and if they helped in prioritizing important concepts.	• The remedial recommendations matched the areas where I needed improvement. • The suggestions were relevant to my recent mistakes. • The recommendations helped me focus on important concepts.
Section B: Clarity & Quality-assessed how well the recommendations were written and structured, and whether the practice problems provided were suitable in terms of difficulty.	• The recommendations were clearly written and easy to understand. • The structure (learning objectives, topics, explanation, and practice) made the recommendations useful. • The practice problems were well-designed and appropriate in difficulty.
Section C: Perceived Impact on Learning-aimed to capture how the recommendations influenced the learners' confidence and conceptual clarity.	• After following the recommendations, I felt more confident about the concept. • The recommendations helped me see what I was missing.
Section D: Engagement & Satisfaction-measured how engaging and motivating the recommendations were and whether learners preferred personalized AI-driven feedback over generic suggestions.	• I found the recommendations engaging and motivating. • I prefer having these AI-generated personalized recommendations over only generic feedback. • Overall, I am satisfied with the personalized recommendation system.
Section E: Open-ended Feedback-collected qualitative feedback to understand what aspects learners found most helpful and to gather suggestions for improvement.	• What did you find most helpful about the recommendations? • What improvements would you suggest for the recommendations? • Any additional comments or feedback?

D. DISCUSSION

This study shows that AI-driven solutions could improve individualized learning and evaluation procedures. The use of generative AI models to make multiple-choice questions that were very similar in meaning to questions made by teachers across a range of topics shows how well Retrieval-Augmented Generation (RAG) strategies work for making assessments that are appropriate for teaching. Among the models tested, the questions made using OpenAI were always more similar to human-made material, which is why it was chosen for primary MCQ generating. The use of Exercise-Aware Knowledge Tracing (EKT) made the system even more adaptable. The trained EKT model was able to predict learners' changing knowledge states better than other models like DKVMN, AKT, and SimpleKT. EKT's dynamic monitoring of concept mastery made it possible to quickly find knowledge gaps and create personalized recommendations for how to fix them. The introduction of personalized remedial recommendations based on individual knowledge gaps led to quantifiable enhancements in learner performance. Comparative analyses conducted pre- and post-intervention demonstrated a significant transition from lower to higher competency levels in foundational computer science concepts. These findings show the value of combining AI-driven assessment, knowledge tracing, and targeted, concept-specific feedback to support learner development and enhance educational outcomes.

To evaluate the effectiveness of the overall process, we conducted a survey among the learners. This survey was designed to evaluate the effectiveness and user satisfaction of an AI-generated personalized recommendation system provided to learners after a skill assessment. The questions were grouped into five key sections as given in Table 7

The bar chart (Figure 9) illustrates the average agreement levels (on a 1–5 Likert scale) for each survey question, offering insights into how learners perceived the AI-generated personalized recommendation system. The highest-scoring items, with averages above 3.3, relate to the relevance and structure of the recommendations—learners felt the system helped them focus on key concepts and found the content clearly written and satisfying overall. Mid-range scores (around 3.0 to 3.3) were observed for questions addressing confidence, motivation, and the ability of the recommendations to help learners identify gaps in their understanding. The lowest-scoring items, averaging below 3.0, pertained to the system's ability to personalize feedback to individual mistakes and learning needs. This suggests that while the overall structure and clarity of the recommendations were well-received, there is room for improvement in making the personalization aspects more accurate and transparent to learners.

The results not only validate the effectiveness of the proposed modules but also highlight the uniqueness of our integrated strategy. Unlike prior works that treat adaptive assessment, MCQ generation, or recommendation as isolated tasks, our system establishes a closed learning loop that connects EKT-based knowledge tracing, RAG-driven assessment, and multi-model cross-verified remedial recommendations within an OBE framework. This seamless integration ensures both conceptual soundness and pedagogical relevance, addressing limitations in existing ensemble or validation approaches. While the framework builds upon

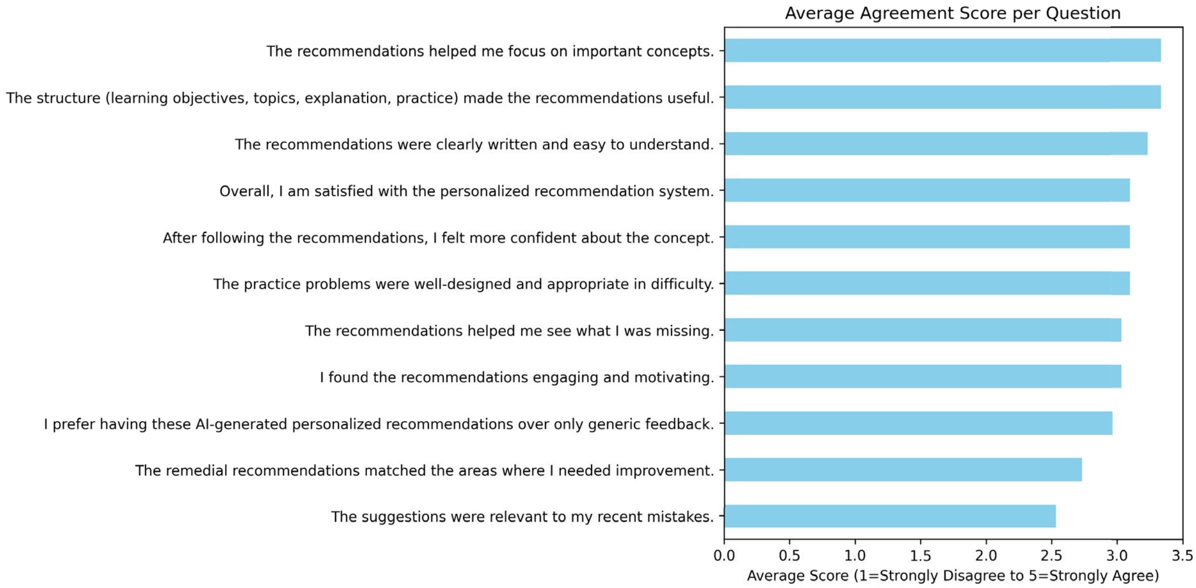

FIGURE 9. Learners' average agreement on the effectiveness of AI-generated recommendations.

established AI models, its novelty lies in combining these methods into a single adaptive learning ecosystem validated in a real-world institutional setting. This applied contribution demonstrates a clear scientific and practical step forward, offering a pathway for transitioning from theoretical AI models to functional, ethically responsible educational tools.

E. ETHICAL AND PRACTICAL CONSIDERATIONS FOR AI-DRIVEN EDUCATIONAL ASSESSMENT

A critical aspect of deploying AI-driven assessment in education is ensuring fairness, transparency, and trustworthiness. To reduce potential algorithmic bias in both MCQ generation and remedial recommendations, the system restricts its knowledge base to selected, curriculum-aligned materials and applies multi-model cross-verification to detect and limit topic drift or unbalanced content. Teachers are still involved in reviewing flagged outputs, especially for tests with high stakes. This adds another layer of protection against unintentional bias. To promote transparency, each remedial suggestion is accompanied by a short explanation that highlights the specific concept gaps identified by the knowledge-tracing model, helping learners understand why particular activities are recommended. Data privacy is preserved by anonymizing learner records before analysis, applying role-based access control to stored data, and complying with institutional data-protection guidelines. These measures support the manuscript's stated alignment with SDG 4 by emphasizing fair, explainable, and responsible use of AI in educational settings.

F. EDUCATIONAL IMPACT AND ALIGNMENT WITH SDG 4

In addition to direct academic benefits, this research furthers the overall objectives of Sustainable Development Goal 4 (SDG 4), prioritizing inclusive and equitable quality education. By mapping assessment and remediation strategies against specific learner needs, the model in question enables ongoing learning, enhances equitable access to resources for learning, and encourages competency-based learning environments. Also, through its compliance with Outcome-Based Education (OBE) standards, it ensures that evaluations of learning remain aligned with practical skills and competencies.

Overall, the findings affirm the feasibility and educational value of AI-driven personalized learning systems, offering a scalable approach to enhancing academic engagement, addressing knowledge gaps, and promoting lifelong learning pathways.

VI. CONCLUSION AND FUTURE DIRECTIONS

This study demonstrated the effectiveness of AI-powered tools in enhancing assessment and remediation processes within educational contexts. By leveraging generative AI models, we were able to generate objective-type questions, evaluate learner responses, and identify conceptual knowledge gaps. The system further provided automated, structured remedial recommendations to help learners strengthen their understanding. These recommendations were categorized into learning objectives, review topics, and practice activities—creating a targeted and efficient learning loop. The evaluation results showed strong alignment between AI-generated output and teacher expectations, indicating the practical utility of the approach. Hybrid profiling methods combining behavioral analysis with assessment results enhance the accuracy of learner classification in adaptive systems, ensuring tailored interventions for diverse needs.

Our evaluation was conducted within a Computer Science program at a single institution, which may restrict generalizability. However, the framework can be easily adapted with different academic structures, such as those that follow Bloom's Taxonomy, OBE, or ABET standards, by using a Retrieval-Augmented Generation (RAG) pipeline that works with standard PDF curricula and assessment data. Additionally, the vector database, recommendation modules, and MCQ generation components are not tied to any one field, so they can be adjusted with knowledge from specific fields to work with a wide range of subjects, such as engineering, business, and health sciences. Future studies will extend evaluation across multiple institutions and domains to establish broader applicability.

This study clearly aligns with the objectives of Sustainable Development Goal 4 (SDG 4) by presenting an AI-assisted framework that improves the quality and inclusivity of education through transparent, interpretable, and learner-centered design. The technology supports fair learning by automatically evaluating learner responses and coming up with useful explanations and individualized suggestions for how to improve. This is especially helpful for learners who are having trouble with certain concepts. The structured feedback, which is broken down into learning objectives, revision topics, remedial explanations, and practice exercises, makes sure that each student gets the right kind of help that they can understand and that is matched to their needs.

The survey questions (Table 7) and the analysis of the feedback (Fig. 9) show that learners are quite accepting of AI-generated recommendations. Participants said that the recommendations were clearer, fairer, and more trustworthy. These results show that the suggested framework not only automates assessment but also makes things more clear, understandable, and engaging for learners, which are all important parts of using AI in education in a way that will last. The framework enhances both the technical and pedagogical aspects of AI-driven personalization by integrating human-in-the-loop mechanisms and explainable feedback generation. This fosters a more inclusive, responsive, and high-quality educational ecosystem, in accordance with the global objectives of SDG 4.

APPENDIX A
ADAPTIVE MCQ GENERATION DETAILED WORKFLOW

The adaptive question generation module uses a Retrieval-Augmented Generation (RAG) approach to create curriculum-aligned multiple-choice questions that vary in difficulty and remain within the scope of instructor-defined concepts. For each knowledge concept, the system generates approximately three times the required number of questions to ensure sufficient variability. Curriculum documents are processed with RAG to extract learning objectives, and all generated questions are constrained to this extracted knowledge base to avoid out-of-context content. For example, if a concept requires five questions per level (easy, medium, hard), about fifteen questions are generated for each level, maintaining both diversity and balanced difficulty.

To support efficient retrieval and adaptive questioning, the text from curriculum PDFs is extracted using `PyPDFLoader`, segmented into overlapping character-based chunks (about 1000 characters with a 100-character overlap), and stored as semantic embeddings in a FAISS vector database. A conversational retrieval chain—integrating a language model such as GPT-4 with the vector store—enables context-aware querying across multi-turn sessions. In addition, a global memory maintains all previously generated questions as semantic vectors to avoid duplication and improve continuity. This combination of controlled retrieval, persistent memory, and Bloom's-aligned difficulty modeling ensures fair, reliable, and concept-focused adaptive assessments.

APPENDIX B
DETAILED ALGORITHM FOR THE PROPOSED METHODOLOGY

Algorithm 1 outlines the overall workflow to generate personalized learning recommendations and adaptive evaluations using curriculum content, OBE evaluation results, and generative AI models.

At the beginning of an academic year, the proposed system will be fine-tuned with the academic program curriculum. The process begins with curriculum processing (Step 1). A Retrieval-Augmented Generation (RAG) model is applied to convert the course curriculum into vector embeddings, enabling semantic search and concept retrieval throughout the process. In Step 2, the learner knowledge gap is identified; for that, each exam question answered by the learner is mapped to the corresponding concepts. Performance data, including concept-level attainment (1-pass/0-fail), is input into an Exercise-Aware Knowledge Tracing (EKT) model. This model generates a latent knowledge state that represents the learner's current understanding across all tracked concepts. By comparing this state against predefined Outcome-Based Education (OBE) thresholds, the system identifies weak or non-attained concepts.

In Step 3, the system generates personalized remedial recommendations for the flagged concepts using the OpenAI generative AI model. These outputs are then cross-verified using secondary models such as Gemini or DeepSeek to ensure alignment, relevance, and quality. The refined recommendations—consisting of explanations, structured breakdowns, and practice tasks—are delivered directly to the learner.

Step 4 focuses on multiple-choice question generation. For each non-attained concept, the system generates concept-aligned multiple-choice questions (MCQs) using the RAG model. Semantic similarity checks (e.g., via S-BERT) are used to filter out duplicates. Further validation ensures each question aligns with the appropriate Bloom's Taxonomy level and maintains conceptual relevance.

In Step 5, the system builds a personalized evaluation set. This set includes questions tailored to the learner's weak concepts and adheres to instructor-defined rules regarding structure and difficulty. After the learner completes the personalized evaluation, the learner's results are reintroduced into the knowledge tracing module, allowing the cycle to continue iteratively until mastery is achieved.

This end-to-end algorithm not only personalizes learning interventions but also ensures continuous adaptation based on performance feedback, making the system suitable for scalable, AI-driven educational environments.

REFERENCES

[1] J. C. Nesbit and O. O. Adesope, "Learning with concept and knowledge maps: A meta-analysis," *Rev. Educ. Res.*, vol. 76, no. 3, pp. 413–448, Sep. 2006.

[2] J. W. Berry and S. L. Chew, "Improving learning through interventions of student-generated questions and concept maps," *Teaching Psychol.*, vol. 35, no. 4, pp. 305–312, Oct. 2008.

[3] D. C. West, J. R. Pomeroy, J. K. Park, E. A. Gerstenberger, and J. Sandoval, "Critical thinking in graduate medical education: A role for concept mapping assessment?" *Jama*, vol. 284, no. 9, pp. 1105–1110, 2000.

[4] C. T. Machado and A. A. Carvalho, "Concept mapping: Benefits and challenges in higher education," *J. Continuing Higher Educ.*, vol. 68, no. 1, pp. 38–53, Jan. 2020.

[5] A. H. Sapci and H. A. Sapci, "Artificial intelligence education and tools for medical and health informatics students: Systematic review," *JMIR Med. Educ.*, vol. 6, no. 1, Jun. 2020, Art. no. e19285.

[6] S. Hubalovsky, M. Hubalovska, and M. Musilek, "Assessment of the influence of adaptive E-learning on learning effectiveness of primary school pupils," *Comput. Hum. Behav.*, vol. 92, pp. 691–705, Mar. 2019.

[7] C. Merino-Campos, "The impact of artificial intelligence on personalized learning in higher education: A systematic review," *Trends Higher Educ.*, vol. 4, no. 2, p. 17, Mar. 2025.

[8] A. Bozkurt, A. Karadeniz, D. Baneres, A. E. Guerrero-Roldán, and M. E. Rodríguez, "Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century," *Sustainability*, vol. 13, no. 2, p. 800, Jan. 2021.

[9] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "E-learning: Challenges and research opportunities using machine learning & data analytics," *IEEE Access*, vol. 6, pp. 39117–39138, 2018.

[10] N. Barkoczi, M. L. Maier, and A. Horvat-Marc, "The impact of artificial intelligence on personalized learning in stem education," in *Proc. INTED*, vol. 1, Mar. 2024, pp. 4980–4989.

[11] R. M. Harden, "Developments in outcome-based education," *Med. Teacher*, vol. 24, no. 2, pp. 117–120, Jan. 2002.

[12] M. Ovinis, S. Karuppanan, S. A. Sulaiman, P. S. Melor, M. Z. Paiz, and A. Urquía, "A comparative analysis of attainment of program outcomes for courses with and without the use of modern tools," *Proc. MATEC Web Conf.*, vol. 225, Jan. 2018, Art. no. 06022.

[13] B. M. Hejazi, "Outcomes-based education (OBE): A transformational perspective on quality and mobility in higher education," *Community College Leadership Program*, pp. 1–30, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:112313343

[14] H. M. Asim, A. Vaz, A. Ahmed, and S. Sadiq, "A review on outcome based education and factors that impact student learning outcomes in tertiary education system," *Int. Educ. Stud.*, vol. 14, no. 2, pp. 1–11, Jan. 2021.

[15] W. Fan, Y. Ding, L. Ning, S. Wang, H. Li, D. Yin, T.-S. Chua, and Q. Li, "A survey on RAG meeting LLMs: Towards retrieval-augmented large language models," in *Proc. 30th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2024, pp. 6491–6501.

[16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 9459–9474.

[17] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani, and M. Usman, "AI-based personalized E-learning systems: Issues, challenges, and solutions," *IEEE Access*, vol. 10, pp. 81323–81342, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9856321

[18] Q. Liu, Z. Huang, Y. Yin, E. Chen, H. Xiong, Y. Su, and G. Hu, "EKT: Exercise-aware knowledge tracing for student performance prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 1, pp. 100–115, Jan. 2021.

[19] R. Sajja, Y. Sermet, M. Cikmaz, D. Cwiertny, and I. Demir, "Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education," *Information*, vol. 15, no. 10, p. 596, Sep. 2024.

[20] N. Tang, C. Yang, J. Fan, L. Cao, Y. Luo, and A. Halevy, "VerifAI: Verified generative AI," 2023, *arXiv:2307.02796*.

[21] S. Webb, J. Holford, S. Hodge, M. Milana, and R. Waller, "Lifelong learning for quality education: Exploring the neglected aspect of sustainable development goal 4," *Int. J. Lifelong Education*, vol. 36, no. 5, pp. 509–511, 2017.

[22] M. Taşkın, "Artificial intelligence in personalized education: Enhancing learning outcomes through adaptive technologies and data-driven insights," *Hum. Comput. Interact.*, vol. 8, no. 1, p. 173, Jan. 2025. [Online]. Available: https://globalresearchandinnovationpublications.com/HCI/article/view/134

[23] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10124747

[24] A. Kumar, A. Kaur, M. Pathania, and A. K. Singh, "MCQGen: A large language model-driven MCQ generator for personalized learning," *IEEE Access*, vol. 12, pp. 102261–102273, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10577164

[25] I. Celik, E. Gedrimiene, S. Siklander, and H. Muukkonen, "The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education," *Australas. J. Educ. Technol.*, vol. 40, no. 3, pp. 19–38, Jun. 2024. [Online]. Available: https://ajet.org.au/index.php/AJET/article/view/9069

[26] L. Murali, G. Gopakumar, D. M. Viswanathan, R. Raman, and P. Nedungadi, "Integrating LLMs and knowledge graphs for medical AI: Advances, challenges, and future directions," *IEEE J. Biomed. Health Informat.*, early access, Oct. 16, 2025, doi: 10.1109/JBHI.2025.3622058.

[27] H. Peng, S. Ma, and J. M. Spector, "Personalized adaptive learning: An emerging pedagogical approach enabled by a smart learning environment," *Smart Learn. Environ.*, vol. 6, no. 1, pp. 1–14, Dec. 2019. [Online]. Available: https://slejournal.springeropen.com/articles/10.1186/s40561-019-0099-9

[28] B. N. Thanh, D. T. H. Vo, M. N. Nhat, T. T. T. Pham, H. T. Trung, and S. Ha Xuan, "Race with the machines: Assessing the capability of generative AI in solving authentic assessments," *Australas. J. Educ. Technol.*, vol. 39, no. 5, pp. 59–81, Dec. 2023.

[29] S. S. Reddy, N. Girish, A. Kumar B, S. Jahnavi, and C. Nandini, "AI-driven personalized learning and content extraction: An emerging paradigm," *J. Inf. Secur. Syst. Cyber Criminol. Res.*, vol. 1, no. 3, pp. 32–39, 2025. [Online]. Available: https://matjournals.net/engineering/index.php/JoISSCCR/article/view/1249

[30] R. Ejjami, "The adaptive personalization theory of learning: Revolutionizing education with AI," *J. Next-Gener. Res. 5.0*, vol. 1, no. 1, pp. 1–18, 2024. [Online]. Available: https://doi.org/10.70792/jngr5.0.v1i1.8

[31] CloudThat Technologies. (Mar. 2024). *The Ethics of AI: Addressing Bias, Privacy, and Accountability in Machine Learning*. [Online]. Available: https://www.cloudthat.com/resources/blog/the-ethics-of-ai-addressing-bias-privacy-and-accountability-in-machine-learning

[32] A. N. Nampoothiri, M. G. Thushara, A. Santhosh, A. Kumar, and S. Kumar, "The impact of transformer models and regression algorithms on automated short answer grading," in *Proc. 5th Int. Conf. Adv. Electr., Comput., Commun. Sustain. Technol. (ICAECT)*, Jan. 2025, pp. 1–6.

[33] C. A. Eden, O. N. Chisom, and I. S. Adeniyi, "Integrating AI in education: Opportunities, challenges, and ethical considerations," *Magna Scientia Adv. Res. Rev.*, vol. 10, no. 2, pp. 6–13, Mar. 2024. [Online]. Available: https://magnascientiapub.com/journals/msarr/content/integrating-ai-education-opportunities-challenges-and-ethical-considerations

[34] K. I. K. Gyonyoru and J. Katona, "Student perceptions of AI-enhanced adaptive learning systems: A pilot survey," in *Proc. IEEE 7th Int. Conf. Workshop Óbuda Electr. Power Eng. (CANDO-EPE)*, Oct. 2024, pp. 93–98.

[35] K. I. K. Gyonyoru and J. Katona, "Comprehensive overview of the concept and applications of AI-based adaptive learning," *Acta Polytechnica Hungarica*, vol. 22, no. 3, pp. 167–186, 2025.

[36] M. Bond, H. Khosravi, M. De Laat, N. Bergdahl, V. Negrea, E. Oxley, P. Pham, S. W. Chong, and G. Siemens, "A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and Rigour," *Int. J. Educ. Technol. Higher Educ.*, vol. 21, no. 1, p. 4, Jan. 2024.

[37] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?" *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–27, Dec. 2019, doi: 10.1186/s41239-019-0171-0.

[38] N. Pradeesh, T. Remya, M. Thushara, K. A. Krishna, and V. Pranav, "Retrieval-augmented generation for multiple-choice questions and answers generation," *Proc. Comput. Sci.*, vol. 259, pp. 504–511, Jan. 2025.

[39] V. Tam, E. Y. Lam, and S. T. Fung, "A new framework of concept clustering and learning path optimization to develop the next-generation e-learning systems," *J. Comput. Educ.*, vol. 1, no. 4, pp. 335–352, Dec. 2014.

[40] S. Minn, "AI-assisted knowledge assessment techniques for adaptive learning environments," *Comput. Educ., Artif. Intell.*, vol. 3, Jan. 2022, Art. no. 100050.

[41] A. S. Arunachalam and T. Velmurugan, "Measures for predicting success factors of e-learning in educational institutions," *Int. J. Pure Appl. Math.*, vol. 118, no. 18, p. 3673, 2018.

[42] S. M. D. and S. Krishnamoorthy, "Knowledge graphs for representing knowledge progression of students across heterogeneous learning systems," *Int. J. Artif. Intell. Educ.*, pp. 1–29, 2025. [Online]. Available: https://doi.org/10.1007/s40593-024-00434-w

[43] M. Soumya and S. Krishnamoorthy, "Student performance prediction, risk analysis, and feedback based on context-bound cognitive skill scores," *Educ. Inf. Technol.*, vol. 27, no. 3, pp. 3981–4005, Apr. 2022.

[44] R. Parvathy, M. G. Thushara, and J. M. Kannimoola, "Automated code assessment and feedback: A comprehensive model for improved programming education," *IEEE Access*, vol. 13, pp. 56642–56658, 2025.

[45] N. S. Raj and V. G. Renumol, "A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020," *J. Comput. Educ.*, vol. 9, no. 1, pp. 113–148, Mar. 2022.

[46] A. Klašnja-Milićević, M. Ivanović, B. Vesin, and Z. Budimac, "Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques," *Appl. Intell.*, vol. 48, no. 6, pp. 1519–1535, Jun. 2018.

[47] M. Riyahi and M. K. Sohrabi, "Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity," *Electron. Commerce Res. Appl.*, vol. 40, Mar. 2020, Art. no. 100938.

[48] X. Ren, W. Yang, X. Jiang, G. Jin, and Y. Yu, "A deep learning framework for multimodal course recommendation based on LSTM+attention," *Sustainability*, vol. 14, no. 5, p. 2907, Mar. 2022.

[49] Y. Zhang, C. Zhao, M. Yuan, M. Chen, and X. Liu, "Unifying attentive sparse autoencoder with neural collaborative filtering for recommendation," *Intell. Data Anal.*, vol. 26, no. 4, pp. 841–857, Jul. 2022.

[50] S. Pandey and J. Srivastava, "RKT: Relation-aware self-attention for knowledge tracing," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 1205–1214.

[51] X. Wang, Y. Li, and Z. Zhang, "Reinforcement learning-based personalized recommendation system," *J. Intell. Syst.*, vol. 29, no. 1, pp. 123–135, 2020.

[52] R. Tertulino and R. Almeida, "Privacy-preserving personalization in education: A federated recommender system for student performance prediction," 2025, *arXiv:2509.10516*.

[53] P. H. Cher, J. W. Y. Lee, and F. Bello, "Machine learning techniques to evaluate lesson objectives," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Switzerland: Springer, 2022, pp. 193–205.

[54] J. Chen, H. Lu, H. Zhou, and Y. Zhou, "Exploration on curriculum teaching based on OBE and AI," in *Proc. 10th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Aug. 2019, pp. 385–389.

[55] S. Keskın and H. Yurdugül, "E-learning experience: Modeling students' e-learning interactions using log data," *J. Educ. Technol. Online Learn.*, vol. 5, no. 1, pp. 1–13, Jan. 2022.

[56] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*. Cham, Switzerland: Springer, 2007, pp. 3–53.

[57] J. D. Novak and A. J. Cañas, "The theory underlying concept maps and how to construct and use them," in *Technical Report IHMC CmapTools*. Institute for Human and Machine Cognition, 2008.

[58] S. Chookaew, P. Panjaburee, D. Wanichsan, and P. Laosinchai, "A personalized E-learning environment to promote student's conceptual learning on basic computer programming," *Proc. Social Behav. Sci.*, vol. 116, pp. 815–819, Feb. 2014.

[59] J. Biggs, "Enhancing teaching through constructive alignment," *Higher Educ.*, vol. 32, no. 3, pp. 347–364, Oct. 1996.

[60] W. Gan and Y. Sun, "Knowledge interaction enhanced knowledge tracing for learner performance prediction," *IEEE Access*, vol. 7, pp. 174430–174442, 2019.

[61] A. Ghosh, N. Heffernan, and A. S. Lan, "Context-aware attentive knowledge tracing," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 2330–2339.

[62] C. Piech, J. Spencer, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," 2015, *arXiv:1506.05908*.

[63] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," 2016, *arXiv:1611.08108*.

[64] P. Kulkarni and P. Joshi, *Artificial Intelligence: Building Intelligent Systems*. New Delhi, India: PHI Learning Pvt. Ltd., 2015.

[65] J. Lee, X. Wang, D. Schuurmans, M. Bosma, E. H., Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 24824–24837.

[66] Y. B. David, A. Segal, and Y. Gal, "Sequencing educational content in classrooms using Bayesian knowledge tracing," in *Proc. 6th Int. Conf. Learn. Anal. Knowl. (LAK)*, 2016, pp. 354–363.

[67] Amrita Vishwa Vidyapeetham. (2020). *Amrita's Online Learning Platform Ample*. Accessed: Apr. 10, 2025. [Online]. Available: https://www.amrita.edu/news/amritas-online-learning-platform-ample/

[68] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, Vancouver, BC, Canada, 2019.

[69] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 3111–3119.

[70] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," 2019, *arXiv:1908.10084*.

[71] R. Řehřek and P. Sojka. (2011). *Gensim-Statistical Semantics in Python*. [Online]. Available: genism.org

[72] C. N. Hang, C. Wei Tan, and P.-D. Yu, "MCQGen: A large language model-driven MCQ generator for personalized learning," *IEEE Access*, vol. 12, pp. 102261–102273, 2024.

[73] K. Vachev, M. Hardalov, G. Karadzhov, G. Georgiev, I. Koychev, and P. Nakov, "Generating answer candidates for quizzes and answer-aware question generators," 2021, *arXiv:2108.12898*.

[74] J. P. Kincaid, R. P. Fishburne Jr., R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," Naval Tech. Training Command Millington TN Res. Branch, 1975.

[75] M. Post, "A call for clarity in reporting BLEU scores," 2018, *arXiv:1804.08771*.

[76] Z. Liu, Q. Liu, J. Chen, S. Huang, and W. Luo, "SimpleKT: A simple but tough-to-beat baseline for knowledge tracing," 2023, *arXiv:2302.06881*.

N. PRADEESH received the degree in computer science from Amrita Vishwa Vidyapeetham, Amritapuri. He is currently a Research Scholar with the School of Computing, Amrita Vishwa Vidyapeetham. He also works as the Project Manager with Omnex Software Solutions Pvt. Ltd., and involves in the development of a learning management systems called AMPLE, which is used in multiple universities, including Amrita Vishwa Vidyapeetham. He has over 14 years of software development experience in various companies. His research interests include education technology, learning management systems, and knowledge graph.

M. G. THUSHARA received the Ph.D. degree in computer science and engineering from Amrita Vishwa Vidyapeetham, Coimbatore, India, in 2021.

In 2010, she was selected for the Erasmus Mundus Scholarship for research at the University of Munich, Germany, for a period of 36 months. In 2012 and 2013, she was invited to work for one month at the DALI Laboratory, University of Perpignan, France, in the domain of abstract interpretation models. Since 2013, she has been an Assistant Professor with the Amrita School of Computing, Amrita Vishwa Vidyapeetham. She works in the domain of program analysis, static analysis for floating point programs, software engineering, automatic code generation and assessment models, and large language models.

K. ARUN KRISHNA He is a studious and enthusiastic learner, actively engaged in AI-driven projects, and is recognized for his strong programming skills.

V. PRANAV is currently pursuing the degree with the School of Computing, Amrita Vishwa Vidyapeetham, specializing in machine learning and artificial intelligence. He is passionate about emerging AI technologies, consistently involved in innovative projects, and demonstrates strong proficiency in programming.

SHIVSUBRAMANI KRISHNAMOORTHY received the Ph.D. degree in computer science from the University of Maryland.

He has research and teaching experience for more than 17 years in various disciplines of computer science. He has more than 30 publications in reputed journals, international conferences, and other venues, with his work being published by IEEE, ACM, Elsevier, and Springer. His research interests include pervasive computing, context-aware computing, mobile computing, and system integration. A public safety system championed by him during his Ph.D. degree at the University of Maryland was widely recognized by prominent news agencies, like CNN, BBC, Al-Jazeera, Fox News, ABC, NBC, NPR, and others covering it. The project was also recognized at various venues, with it receiving awards like the VITA-Wireless Samarian Award by the CTIA Wireless Association and the National Technology Champion of the Year by the National Organization of Black Law Enforcement (NOBLE), USA.