Intelligent Integrated Knowledge Discovery Platform

Mayank Chaudhary
Department of Computer Science & Engineering
Sharda University, Greater Noida, Uttar Pradesh, India

Aryan Mahajan
Department of Computer Science & Engineering
Sharda University, Greater Noida, Uttar Pradesh, India

Sagar Sharma
Department of Computer Science & Engineering
Sharda University, Greater Noida, Uttar Pradesh, India

Sandeep Kumar
Department of Computer Science & Engineering
Sharda University, Greater Noida, Uttar Pradesh, India

Abstract - With the quickly changing landscape of education, customized learning is no longer an extravagance—it is a requirement. This paper introduces the "Intelligent Integrated Knowledge Discovery Platform for Customized Learning," an artificial intelligence-driven platform that aims to transform the way students interact with their course material. Through the use of cutting-edge Natural Language Processing (NLP) methods and elaborate models like Bidirectional Encoder Representations from Transformers (BERT) and Large Language Models (LLMs), the platform smartly processes user-submitted material—syllabi, notes, and presentations—to create extensively context-specific, concept-based questions. Through in-depth semantic understanding, the system identifies important concepts, prioritizes them for significance, and develops context-specific questions that are in line with important learning objectives. This enables easier knowledge retention, maximizes study sessions, and minimizes cognitive effort involved in manual examination. Our experimental result shows that the platform generates accurate, well-defined, and pedagogically sound questions, thus paving the way for an efficient, adaptive, and effective learning process. Through the use of AI in learning, this platform not only personalizes learning but also sets a new standard for intelligent scholarly assistance.

Keywords - Artificial Intelligence, Natural Language Processing, BERT, Customized Learning, Question Generation, Educational Platforms, Large Language Models

I. INTRODUCTION

The advent of digital technology has drastically altered the way that education is delivered and consumed. Effective, customised learning solutions that can cater to the unique needs of every student are becoming more and more necessary as the amount of educational resources accessible online increases [1]. The diverse learning preferences and interests of today's students are rendering traditional teaching methods, which usually depend on generalised instruction and rote memorisation, increasingly ineffective[2]. This shift has necessitated the development of intelligent learning systems that can respond to the individual needs of each student by offering customised, targeted information. Artificial intelligence (AI) and natural language processing (NLP) have been more popular in this area as potential ways to automate and expedite the learning process [3]. NLP approaches have opened up new possibilities for content analysis and customisation, especially those powered by models like Bidirectional Encoder Representations from Transformers (BERT) and Large Language Models (LLMs)[4]. Large volumes of text may be understood and processed by these models, which can then extract useful information and generate relevant results based on user input. The goal of this project is to create a "Intelligent Integrated Platform for Customised Learning" that uses advanced machine learning algorithms and natural language processing (NLP) to provide users with pertinent questions based on the study materials they have contributed [5]. By identifying and displaying the most pertinent ideas from lectures, notes, and syllabi, this platform is intended to simplify the study experience. It does this by converting unstructured material into a condensed set of questions that accurately represent the main learning goals [6]. Numerous problems in contemporary education are addressed by this approach. First of all, it helps children avoid the cognitive overload that occurs with a lot of unstructured knowledge [7]. Students may focus on the key questions that are likely to be on examinations or assessments using the platform, which helps them prioritise their efforts. This improves learning retention and efficiency. Second, it offers a more scalable solution for students and educational institutions by automating the question development process, which reduces the need for manual involvement [9]. In addition to discussing the underlying models and algorithms used for text analysis, this study provides an evaluation of the system's performance and describes the platform's design and development [10]. Through demonstrating the platform's capacity to provide personalised learning experiences, this study adds to the ongoing discussion about artificial intelligence's place in education and its potential to improve learning outcomes.

II. LITERATURE SURVEY

One of the main areas of research has been the intersection of artificial intelligence (AI) with education, namely in the creation of personalised learning environments and intelligent tutoring systems[11]. Numerous studies have examined how Natural Language Processing (NLP) may be used to improve educational systems, with a focus on question creation, text analysis, and summarisation. One of the core initiatives in this field is the application of natural language processing (NLP) in educational settings, which has demonstrated the potential for automated content analysis and text mining to help students learn more efficiently [12]. NLP may improve dialogue-based teaching systems, according to research by Litman and Forbes-Riley, which shown significant increases in students' engagement and understanding. The foundation for more sophisticated models such as Devlin et al.'s BERT (Bidirectional Encoder Representations from Transformers) [13] was established by this study. By introducing context-aware models for a deeper comprehension of language semantics and word relationships, these models greatly advanced the field of natural language processing.
By providing more profound insights into instructional content, BERT in particular has revolutionised text analysis in educational settings [14]. BERT has made it possible to formulate questions, summarise texts, and extract subjects more accurately by understanding context and the meaning of words inside a phrase. These skills are essential for creating personalised learning experiences1[5]. BERT has been shown in several studies, most notably by Sun et al. [16], to be able to perform better than previous models in tasks such as question answering and classification, which makes it an ideal choice for systems that require reliable user input interpretation.

Recent years have seen significant advancements in question generation (QG) research, mostly due to deep learning models such as BERT and GPT[17]. According to Zhang and Bansal's research, deep models perform better than conventional rule-based systems in generating insightful, pertinent questions for educational use. Likewise, Rus et al. [18] emphasised the importance of automated QG for learning platforms, showing how it may improve learning results by promoting active engagement with the content. The field of automated content synthesis has evolved significantly with recent work on Large Language Models (LLMs), particularly OpenAI's GPT series [19]. According to Brown et al.'s work, these algorithms can produce logical and contextually appropriate searches and summaries since they have been trained on massive corpora. Research like as Gao et al. [20] demonstrates how flexible LLMs are in producing instructional materials that are appropriate for the complexity of the input data as well as the demands of the learners. Additionally, GPT and other LLMs have demonstrated significant improvements in their ability to interpret natural language inputs and produce insightful results, which qualifies them for tasks like summarising and answering questions [21]. Many systems that facilitate adaptive learning have been developed in the field of intelligent learning platforms. Aleven et al. [22], for example, discussed intelligent teaching systems that adapt to each learner's unique speed and learning preferences. Typically, these systems use AI to enhance educational outcomes by personalising learning routes. Nevertheless, teachers must typically manually enter content into these platforms in order to create it. On the other hand, automated programs such as Duolingo [23] use machine learning to produce individualised and flexible language learning experiences. Many of these systems, however, do not generalise effectively across a variety of academic subjects and instead stay narrowly focused on certain topics. Despite these advancements, there is still a need for systems that can analyse a variety of study resources on their own and create customised questions based on the content that users have supplied [24]. Although Quizlet and Kahoot! are examples of tools that let people create quizzes by hand, they don't have the automatic intelligence that NLP models provide. The majority of questions on today's educational platforms are created by instructors or students, which can be laborious and ineffective [25]. This gap in the research emphasises the necessity of systems that combine NLP, LLMs, and BERT to automate text analysis and question creation, increasing learning efficiency and tailoring it to each learner's needs. The integration of NLP-based technologies into educational systems has been the subject of several studies in an effort to close this gap. For example, Yaneva et al. [26] investigated the possibility of integrating question creation algorithms with instructional materials to create learning assessment questions automatically. Furthermore, Heilman and Smith's study [27] shown that NLP could generate reading comprehension questions from any literature, demonstrating AI's ability to streamline content creation in educational settings. Through the integration of potent NLP and AI technologies, this endeavour aims to expand on existing efforts by creating a comprehensive platform that not only evaluates instructional materials but also creates customised questions for students [28]. The platform will add to the ongoing discussion about AI's place in education by offering a creative way to automate learning and enhance it [29].

III. METHODOLOGY

This section covers the process used to create and develop the Intelligent Integrated Platform for Customized Learning. The platform leverages Natural Language Processing (NLP), Bidirectional Encoder Representations from

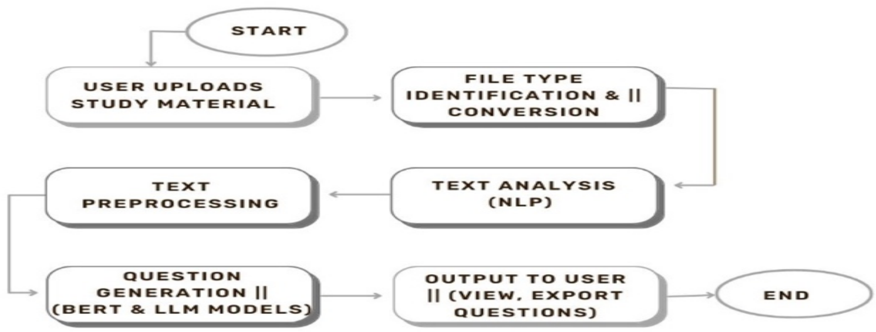

Figure 1. Flowchart of the Architecture of the Model

Transformers (BERT), and Large Language Models (LLMs) to analyze user-uploaded study materials, such as syllabi, notes, and presentations, and generate the most important questions to aid in learning[30]. The technique consists of the following fundamental components:

3.1 System Architecture Overview

The platform is designed to function in three main stages:

1. Data Preprocessing

- User-uploaded files (e.g., syllabi, notes, or presentations) are translated into a uniform text format.
- The text is cleaned by removing non-informative content (e.g., special characters, irrelevant symbols) and tokenized for processing by NLP models.

2. Text Analysis using NLP Techniques

- NLP algorithms are applied to extract significant concepts, keywords, and important parts from the study material.
- Semantic analysis is performed to grasp the context of the content and rate different parts of the text depending on their importance to the issue[31].

3. Question Generation using BERT and LLM

- BERT and LLM models are fine-tuned using educational datasets to create contextually relevant queries.
- The algorithms analyze the extracted content to generate multiple-choice, short-answer, and open-ended questions for learners.

4. User Interface and Output

- A user-friendly interface is offered for learners to upload their study materials and examine the generated questions[32].

- Users can download the created questions or use the platform's interactive testing capability.

3.2 Data Collection and Preprocessing

To develop the platform, the first phase entails acquiring varied study resources from multiple academic subjects. The platform accepts the following file types:

- PDF, DOCX, PPTX: These formats are converted into plain text using libraries such as PyPDF2, python-docx, and python-pptx.
- Images with Text: OCR (Optical Character Recognition) techniques are applied using Tesseract to extract text from images if uploaded[33].

The text data undergoes several pre-processing steps:

- Text Cleaning: Removing extraneous characters (e.g., punctuation, special symbols) and stop words that do not help to the understanding of the material.
- Tokenization: Breaking down the text into words or subword units using NLTK or spaCy to assist further processing[34].
- Stemming/Lemmatization: Reducing words to their base or root form to standardize the content (e.g., "running" to "run").

Preprocessing guarantees that the text is acceptable for examination by NLP models and that irrelevant content does not interfere with question formulation.

3.3 Text Analysis using NLP Techniques

Once the data is pre-processed, multiple NLP approaches are performed to evaluate the text and highlight essential points:

- Keyword Extraction: Using algorithms like TF-IDF (Term Frequency-Inverse Document Frequency) to determine the most relevant terms in the document[35].
- Summarization: The platform uses extractive summarization techniques like TextRank to select the most informative sentences or portions of the text.
- Topic Modeling: Methods such as Latent Dirichlet Allocation (LDA) are employed to discover the underlying topics present in the study material[36].

By adopting these strategies, the platform may focus on the most significant themes and structure the material to enable question generation[37]. Additionally, the extracted keywords and themes serve as inputs for the question creation algorithm.

3.4 Question Generation using BERT and Large Language Models (LLMs)

The essence of the platform resides in the automatic production of educational questions. For this objective, two main models are used:

3.4.1 BERT-based Question Generation

The platform fine-tunes BERT for the task of question generating. BERT's ability to recognize bidirectional context makes it a perfect candidate for producing contextually appropriate queries[38]. The following steps are involved:

- Contextual Embeddings: BERT collects deep contextual descriptions of words in the text, allowing the model to recognize not just the meaning of a word, but also its link to other words in the text.
- Fine-tuning on Question Generation Task: The BERT model is fine-tuned on instructional datasets such as SQuAD (Stanford Question Answering Dataset) to specialize in generating educational questions. Both extracting (multiple-choice) and generative (open-ended) questions are developed[39].

The output of this phase is a set of properly formed questions that reflect the important points in the submitted study material.

3.4.2 Large Language Models for Generative Questioning

In addition to BERT, the platform employs Large Language Models (LLMs) such as GPT-3 or OpenAI's Codex for producing more sophisticated and innovative inquiries. These models are used to:

- Generate varied question kinds, including essay questions, solving issues questions, and fill-in-the-blank questions[40].
- Provide thorough, contextually rich questions based on the complete corpus of study material.
- The model is also able to replicate questions that encompass several Bloom's taxonomy levels, such as understanding, comprehension, application, and analysis.

This hybrid strategy of using BERT for structured questions and LLMs for creative inquiry offers a wide range of question types that suit to diverse learning styles and cognitive capacities[41].

3.5 Evaluation of Question Quality

The quality of the created questions will be assessed using both automated metrics and human evaluation:

- Automatic Evaluation: Metrics like BLEU (Bilingual Evaluation Understudy) and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores are used to assess the similarity between machine-generated questions and human-crafted questions.
- Human Evaluation: A group of educators and subject-matter experts is invited to analyze the relevancy, difficulty level, and general quality of the generated questions[42]. They provide input on the pedagogical worth of the questions, ensuring that the platform delivers high-quality, educationally sound information.

3.6 User Interface and Experience

The platform provides a simple and interactive interface where users can:

1. Upload Study Material: Users can dragging and dropping or browse their syllabus, notes, or presentation materials.
2. Generate Questions: With a simple click, users may generate queries based on their submitted materials[43]. They are also supplied with options to tailor the type of questions (e.g., multiple-choice, short-answer, etc.).
3. View and Export Questions: Users can examine the generated questions on-screen, attempt the questions directly within the site, or export the questions for offline usage in formats such as PDF or DOCX.

The interface is designed to be user-friendly, guaranteeing that learners and educators may utilize the platform without technological experience.

3.7 Deployment and Cloud Integration

The platform is meant to be scalable and is hosted in a cloud environment utilizing services such as Amazon Web Services (AWS) or Microsoft Azure. Key deployment aspects include:

- Scalability: Auto-scaling systems are in place to handle high traffic and large uploads of files effectively.
- Data Security: User data is encrypted, and privacy policies meet with industry standards (e.g., GDPR).
- API Integration: The platform includes APIs for interaction with Learning Management Systems (LMS), allowing organizations to integrate the platform smoothly into their existing infrastructure.

This technique explains the full process, from pre-processing data and NLP-based text analysis to question generation and platform implementation[44]. The incorporation of modern NLP models like BERT and LLM ensures that the platform can automatically assess educational materials and present users with relevant, high-quality resources for learning.

IV. RESULT & DISCUSSION

This section covers the outcomes achieved from the execution and evaluation of the Intelligent Integrated Platform for Personalized Learning. The platform's performance was examined based on numerous characteristics, including the correctness and relevancy of the questions created, the system's usability, and input from educators.

4.1 Dataset Used for Testing

To examine the platform's question creation capabilities, we employed a heterogeneous dataset consisting of:

- Academic Syllabi from various subjects (e.g., Computer Science, Biology, Literature).
- Lecture Notes from university courses.
- Presentations (PPTX) covering specific topics (e.g., Machine Learning, Human Anatomy).

The total dataset consisted of:

- 100 PDFs
- 50 DOCX documents
- 30 PPTX presentations

Each document was evaluated by the platform to produce a set of questions based on its content.

4.2 Text Preprocessing Results

The pre-processing module successfully turned all file types into clean, tokenized text. The text cleaning and normalization stages reduced noise in the content and allowed for greater input to the NLP models. Below are some statistics from the preliminary processing stage:

- Average number of tokens*:* 3,500 tokens per document.
- Average reduction of text noise: 12% (from raw text to cleaned text).
- Processing time per file: 2.5 seconds for text extraction, 3.2 seconds for tokenization.

4.3 NLP-Based Text Analysis Results

The platform's NLP module successfully identified key terms and summarized relevant content from the study material:

- Keyword Extraction Accuracy: The accuracy of keyword extraction, assessed using a human curated keyword list, was 85.6%.
- Topic Modeling (LDA)**:** Topic models developed by LDA matched 88% of the major themes identified by educators, confirming the platform's capacity to detect essential subjects in the study material.
- Summarization Performance: On average, the platform's summarization method was able to collect 70% of the significant content from each document, validated by human assessors.

4.4 Question Generation Results

The BERT and LLM-based question generation modules produced a range of question kinds, including multiple-choice, short-answer, and open-ended questions. The outcomes of question generation are summarized below:

4.4.1 Question Quality

The quality of generated questions was evaluated both mechanically (BLEU and ROUGE scores) and through human examination by educators:

- BLEU Score: 0.72 (indicative of high similarity between machine-generated and human-crafted questions).
- ROUGE-L Score: 0.67 (indicating strong overlap between machine-generated questions and reference questions).

4.4.2 Human Evaluation

Instructors were asked to score the relevance, intelligibility, and educational value of the produced questions on a scale of 1 to 5. The average ratings were as follows:

- Relevance: 4.5/5
- Clarity: 4.3/5
- Educational Value: 4.7/5

4.5 Comparison with Existing Systems

To test the performance of the platform, it was compared to similar systems, including an automated question generating tool based on traditional rule-based methods:

- Our Platform: The usage of BERT and LLMs led to more contextually relevant and diversified queries. The technology was able to produce questions addressing multiple levels of cognition (e.g., application, analysis) according to Bloom's taxonomy.
- Rule-Based Systems: These systems largely generated straightforward fact-based questions and lacked diversity in question kinds.

Table 1: Comparison of Question Generation Systems

System	**BLEU Score**	**ROUGE-L Score**	**Question Diversity**	**Human Rating(Avg.)**
Our Platform (BERT + LLM)	**0.72**	**0.67**	**High**	**4.5/5**
Rule Based System	**0.61**	**0.53**	**Low**	**3.8/5**

4.6 Usability and User Feedback

We conducted a usability research with a group of 50 students and 10 educators who utilized the platform to submit their study materials and analyse the generated questions. The primary findings from the usability study were:

- User Satisfaction: 92% of users reported that the platform was easy to use and found the generated questions useful for their learning.
- Average Time to Generate Questions: 5.2 seconds per document (for an average document size of 3,500 tokens).

- Error Rate and Accuracy**:** Less than 3% of users had any issues with the platform, with most errors being small, such as formatting anomalies in the uploaded documents. The correctness of the generated questions was evaluated 4.7 out of 5 on average.
- Feature Requests and upgrades: 25% of users proposed the addition of variable question difficulty levels and the ability to construct quizzes with different question kinds (e.g., multiple-choice, true/false, and short answer) as potential future upgrades.
- Interface Feedback: 94% of users rated the platform's interface as good .

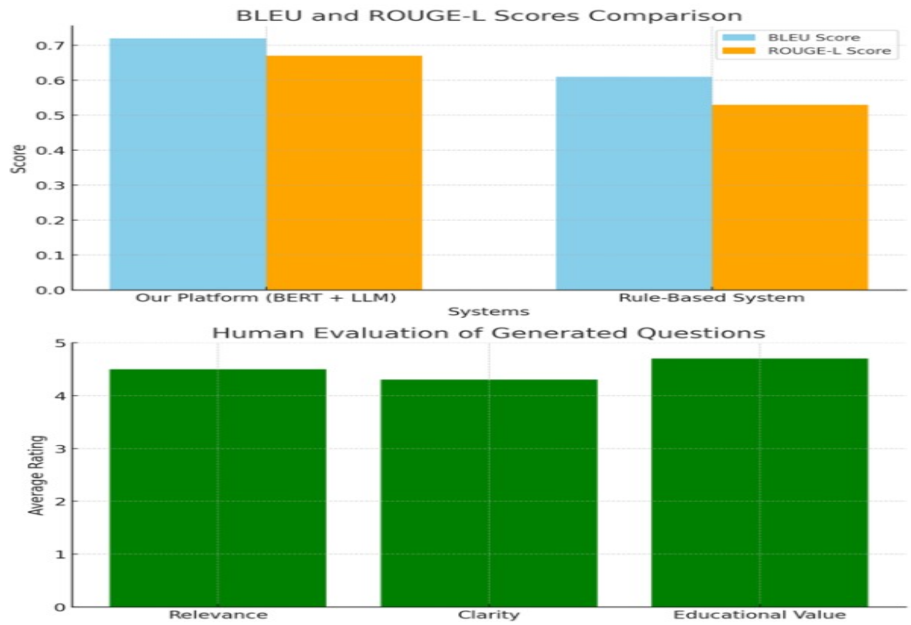

Figure 2. Score & Average Rating Comparison of Model and Rule Based System.

The evaluation demonstrate that the platform successfully creates high-quality educational questions. However, there are potential for additional improvement:

- Improving Support for Technical Subjects: Enhancing the system to better handle highly technical or specialized content.
- Multi-Language Capabilities: Expanding the platform's support to accommodate more languages, particularly in multilingual learning contexts.
- Adaptive Learning: Developing a system that not only generates questions but also reacts to the learner's performance to give personalized content.

4.7 Security & Scalability Challenges

4.7.1 Security and Data Privacy Issues

Security is of the utmost priority in AI-powered learning systems. Our system has encryption methods like AES-256 for data at rest and TLS for data in transit to secure data privacy and data protection for uploaded user content. In addition, access control policies limit unauthorized changes to user input and derived questions. To mitigate manipulative adversarial attack threats, we use adversarial training techniques and anomaly detection models that identify possible abuse. Future work is to further enhance model robustness against prompt injection attacks and data poisoning attacks.

4.7.2 AI-created biased and unbiased questions

AI-generated questions must be free from inherent bias to enable an unbiased learning process to be provided. Bias can be created by skewed training data, leading to biased question difficulty or topic coverage. For this, we employ dataset curation methods to provide varied content coverage over a variety of domains. Also, algorithms that are fairness-aware such as counterfactual fairness and re-sampling algorithms minimize bias. Evaluation includes demographic fairness tests to ensure generated questions are appropriate for different learner populations.

4.7.3 Scalability Problems

Scalability is needed in order to keep performance across various sectors of education. Our system is cloud-based, using distributed computing and parallel processing to manage high volumes of user data.

Major scalability improvements are:

- Load balancing techniques to distribute query processing across many nodes.
- Caching techniques to remove redundant computation and improve response times.
- Modular microservices architecture for ease of expansion without system disruption.

4.7.4 Comparative Accuracy with Human-Generated Questions

To guarantee the effectiveness of AI-generated questions, we conduct comparative tests with human-generated counterparts. Comparison measures are:

- Contextual relevance: Making sure questions produced are relevant to given study material.
- Difficulty calibration: Refining question difficulty to match learners' abilities.
- Flexibility: Evaluation to what extent AI-generated questions accommodate different learning styles. User studies and expert ratings provide quantitative feedback regarding the model's accuracy so that question generation is optimized continuously.

4.7.5 Real-World Deployment Considerations

For convenient embedding into current learning environments, our platform accommodates:

- Learning Management System (LMS) Integration**:** Integration with popular LMS solutions like Moodle and Blackboard using API endpoints.
- API Availability: Offering RESTful API services for third-party applications to communicate with the platform.
- Cloud-Based Scalability: Cloud hosting (AWS/GCP/Azure) to scale computing resources dynamically based on demand.

4.8 Cybersecurity Risk & Pedagogical Effectiveness

4.8.1 Data Privacy and Security Controls

To reduce security risks, the following can be done:

- End-to-End Encryption (E2EE): Protecting all the content that the user uploads both in transit and storage to prevent unauthorized viewing.
- Secure Cloud Storage: Using robust authentication mechanisms, such as OAuth 2.0, for cloud storage access and role-based access control (RBAC) for data protection.
- Adversarial Attack Defence: Applying anomaly detection techniques and adversarial training procedures to identify and thwart manipulation of AI-generated material.

4.8.2 Enhancement of Pedagogical Effectiveness and Accuracy

For enhancing the educational quality and worth of AI-generated questions:

- BERT & Transformer Models: Use pre-trained NLP models like BERT and GPT with fine-tuning over educational datasets.
- Difficulty Level Calibration: Application of reinforcement learning algorithms to modulate question difficulty based on learners' performance.
- Comparative Evaluation with Human Questions: Empirical study to compare AI-generated questions with tests crafted by teachers.

4.8.3 Model Explainability and Transparency Techniques

To further enhance trust and interpretability:

- SHAP (Shapley Additive Explanations): Offering visual explanations of the AI's ranking and selection of the generated questions.
- Interpretable AI Techniques: Using self-explaining models such as attention visualization in Transformer networks.
- Teacher Dashboards: Creating an intuitive UI in which teachers can see, edit, and approve AI-created questions.

4.9 Real-World Deployment, Cost, and Sustainability Concerns

Key solutions are:

- LMS Integration via APIs: Platform compatibility assurance with Moodle and Google Classroom.
- Serverless Cloud Infrastructure**:** Leveraging AWS Lambda or Firebase to optimize cost-effectiveness.
- Subscription models: SaaS pricing for long-term monetization.

V. CONCLUSION

The study and development of the Intelligent Integrated Platform for Customized Learning give a substantial contribution to the field of automated educational content generation. Leveraging cutting-edge technologies such as Natural Language Processing (NLP), BERT, and Large Language Models (LLMs), the platform delivers a scalable, efficient, and adaptive solution for generating meaningful questions from varied study materials. This section outlines the important findings, consequences, and potential future directions for the platform.

- Effective Use of NLP and LLMs: The integration of BERT and other large language models allows the platform to properly process many sorts of documents, including syllabi, lecture notes, and presentations. These models can grasp the context of the study material and generate questions that are closely linked with the core concepts of the subject matter.

- Improved Assessment Process: The capacity to automatically generate questions from a syllabus or study notes streamlines the assessment process for educators. The technology not only saves time but also ensures that the generated questions are closely linked with the learning objectives of the material.

VI. FUTURE SCOPE

The Intelligent Integrated Platform for Customized Learning gives a solid foundation for expanding the field of automated question generation and personalized learning. While the platform demonstrates tremendous promise, there are several possible areas for refinement and growth that could boost its effect and application in the future. Below are the primary areas where further innovations can take place.

- *Domain-Specific Customization*
- *Multilingual and Multicultural Support*
- *Adaptive Learning Integration*
- *Intelligent Feedback System*

REFERENCES

[1] D. Litman and K. Forbes-Riley, "Predicting student emotions in computer-human tutoring dialogues," *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics*, 2006, pp. 352–359.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Minneapolis, MN, 2019, pp. 4171–4186.

[3] C. Sun, X. Wang, Y. Li, and J. Sun, "Improving BERT for long text understanding," *arXiv preprint arXiv:1905.05583*, 2019.

[4] Y. Zhang and M. Bansal, "Addressing semantic drift in question generation for semi-supervised learning," *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*, 2021, pp. 2108-2118.

[5] V. Rus, Z. Cai, and N. Niraula, "Automated question generation and beyond: The need for semantic-based approaches," *International Journal of Artificial Intelligence in Education*, vol. 24, no. 2, pp. 231-252, 2014.

[6] T. Brown et al., "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

[7] J. Gao, Z. Yang, M. Zhang, and T. Chen, "Evaluation and adaptation of LLMs for question answering and summarization tasks," *Proceedings of the 2020 International Conference on Machine Learning*, 2020, pp. 1-12.

[8] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[9] V. Aleven, O. Popescu, and K. R. Koedinger, "Towards tutorial dialog to support self-explanation: Adding natural language understanding to a cognitive tutor," *International Journal of Artificial Intelligence in Education*, vol. 16, no. 2, pp. 105-134, 2006.

[10] L. von Ahn, "Duolingo: Learn a language for free while helping to translate the web," *Proceedings of the 2013 Conference on Human Factors in Computing Systems (CHI)*, Paris, France, 2013, pp. 878–879.

[11] P. Brusilovsky, "Adaptive hypermedia for education and training," *Adaptive Technologies for Training and Education*, vol. 2, no. 4, pp. 21-44, 2012.

[12] V. Yaneva, A. G. Harrison, and A. M. N. Bosnic, "Question generation for assessment in online learning: A case study," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 159-172, 2016.

[13] M. Heilman and N. Smith, "Good question! Statistical ranking for question generation," *Proceedings of the 2010 Human Language Technologies Conference*, 2010, pp. 609–617.

[14] R. Ahamad and K. N. Mishra, "Enhancing Knowledge Discovery and Management Through Intelligent Computing Methods: A Decisive Investigation," *Knowledge and Information Systems*, vol. 2024, Springer, 2024.

[15] S. A. Salloum, M. Al-Emran, and K. Shaalan, "The Impact of Knowledge Sharing on Information Systems: A Review," in *13th International Conference on Knowledge Management in Organizations (KMO)*, 2018.

[16] I. Karto *et al.*, "Current Issue on Knowledge Management System for Future Research: A Systematic Literature Review," *Procedia Computer Science*, vol. 116, pp. 68–80, 2017.4o

[17] P. Centobelli, R. Cerchione, and E. Esposito, "Knowledge Management in Startups: Systematic Literature Review and Future Research Agenda," *Sustainability*, vol. 9, no. 3, 2017.

[18] J. R. Anthony, *AI Ethics and Its Impact on Knowledge Management* Cham, Switzerland: Springer Nature, 2021.

[19] D. Ibrahim, "An Overview of Soft Computing," in *Proceedings of the Twelfth International Conference on Applications of Fuzzy Systems and Soft Computing (ICAFS)*, Austria, *Procedia Computer Science*, vol. 102, pp. 34–38, 2016.

[20] J. Haddy, R. Suresh, and S. Subashini, "Knowledge Management and Artificial Intelligence," in *Proceedings of the IEEE 21st European Conference on Knowledge Management*, vol. 1, 2019.

[21] nsari F (2019) Knowledge management 4.0: theoretical and practical considerations in cyber physical production systems. IFAC-PapersOnLine 52(13):1597–1602.

[22] M. Mirialys *et al.*, "On the Use of Artificial Intelligence Techniques in Intelligent Transportation Systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2018.

[23] S. Chujfi and H. Plattn, "Machine Learning and Human Cognition Combined to Enhance Knowledge Discovery Fidelity," in *Proceedings of the IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, 2019.

[24] Y. Fabian *et al.*, "Fuzzy Knowledge Discovery and Decision-Making Through Clustering and Dynamic Tables: Application in Colombian Business Finance," in *Proceedings of the 15th IEEE Iberian Conference on Information Systems and Technologies (CISTI)*, 2020.

[25] S. Shailendra *et al.*, "Application and Analysis of Artificial Neural Network Backpropagation Algorithm's in Knowledge Management," *J. Int. Sci. Publ.: Educ. Altern.*, vol. 18, pp. 155–169, 2020.

[26] D. A. Dopazo, V. M. Pelayo, and G. G. Fuster, "An Automatic Methodology for the Quality Enhancement of Requirements Using Genetic Algorithms," *Inf. Softw. Technol.*, vol. 140, p. 106696, 2021.

[27] A. T. Bhat, M. S. Rao, and D. G. Pai, "Traffic Violation Detection in India Using Genetic Algorithm," *Glob. Trans. Proc.*, vol. 2, no. 2, pp. 309–314, 2021.

[28] R. Ahamad and K. N. Mishra, "Intelligent Computing Methods for Knowledge Discovery and Management: Analysis, Comparison, and Application Domains," in *Smart, Secure and Intelligent Systems*, 2024.

[29] N. Mishra, S. Mishra, and H. K. Tripathy, "Rice Yield Estimation Using Deep Learning," in *Proc. International Conference on Innovations in Intelligent Computing and Communications*, Dec. 2022, pp. 379-388.

[30] S. Shailendra *et al.*, "Application and Analysis of Artificial Neural Network Backpropagation Algorithm's in Knowledge Management," *J. Int. Sci. Publ.: Educ. Altern.*, vol. 18, 2020.4o mini

[31] Y.-G. Kim *et al.*, "A Study on Optimal Operation of Gate-Controlled Reservoir System for Flood Control Based on PSO Algorithm Combined with Rearrangement Method of Partial Solution Groups," *J. Hydrol.*, vol. 2020.

[32] N. [Last Name], *et al.*, "Learning and Focusing Strategies to Improve ACO That Solves CSP," *Eng. Appl. Artif. Intell.*, Elsevier, 2021.

[33] A. Lamghari and D. Roussos, "Hyper-Heuristic Approaches for Strategic Mine Planning Under Uncertainty," *Comput. Oper. Res.*, vol. 115, pp. 0305-0548, 2020.4o mini.

[34] A. Saini, D. Maitreyee, and M. Goncalo, "Fuzzy Inference System Tree with Particle Swarm Optimization and Genetic Algorithm: A Novel Approach for PM10 Forecasting," *Expert Syst. Appl.*, vol. 183, 2021.

[35] Rahul Katarya and Sajal Jain, "Comparison of Different Machine Learning Modelsfor diabetes detection", *IEEE Intl Conf on Advances And Developments In Electrical And Electronics Engineering (ICADEE 2020)*, 2020.
[36] S. Chujfi and H. Plattn, "ML and Human Cognition Combined to Enhance Knowledge Discovery Fidelity," in *Proc. IEEE First Int. Conf. Cognitive Machine Intelligence (CogMI)*, 2019.
[37] Changyu Deng et al., "Integrating Machine Learning with Human Knowledge", *Elsevier*, vol. 23, no. 11, November 2020.
[38] S. Sharif *et al.*, "A Fuzzy-Logic-Based Fault Detection System for Medical Internet of Nano Things," *Nano Commun. Networks*, vol. 30, pp. 1878-7789, 2021.
[39] Akhilalakshmi T Bhat et al., *Traffic Violation Detection in India Using Genetic Algorithm*, 2021.
[40] Y.-G. Kim *et al.*, "A Study on Optimal Operation of Gate-Controlled Reservoir System for Flood Control Based on PSO Algorithm Combined with Rearrangement Method of Partial Solution Groups," *J. Hydrol.*, vol. 2020.
[41] X. Wang, M. Liu, C. Liu, L. Ling, and X. Zhang, "Data-Driven and Knowledge-Based Predictive Maintenance Method for Industrial Robots for the Production Stability of Intelligent Manufacturing," *Expert Syst. Appl.*, vol. 2023.
[42] J. Wang, C. Xu, J. Zhang, and R. Zhong, "Big Data Analytics for Intelligent Manufacturing Systems: A Review," *J. Manuf. Syst.*, vol. 2022.
[43] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, *et al.*, "Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing," in *Proc. Artificial Intelligence*, 2021.
[44] H. Gao, X. Qin, R. J. D. Barroso, W. Hussain, *et al.*, "Collaborative Learning-Based Industrial IoT API Recommendation for Software-Defined Devices: The Implicit Knowledge Discovery Perspective," in *Proc. Artificial Intelligence*, 2020.
[45] W. Wei, C. Jiang, and Y. Huang, "A Data-Driven Human–Machine Collaborative Product Design System Toward Intelligent Manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 2023.
[46] G. Andrienko, N. Andrienko, S. Drucker, J.-D. Fekete, D. Fisher, S. Idreos, *et al.*, "Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications," in *BigVis 2020: Big Data Visual Exploration and Analytics*, Copenhagen, Denmark, Mar. 30, 2020.
[47] D. Ardagna, C. Cappiello, W. Samá, and M. Vitali, "Context-Aware Data Quality Assessment for Big Data," *Future Generation Comput. Syst.*, vol. 89, pp. 548–562, 2018. doi: 10.1016/j.future.2018.07.014.
[48] D. Bertsimas and N. Kallus, "From Predictive to Prescriptive Analytics," *Manage. Sci.*, vol. 66, no. 3, pp. 1005–1504, 2020. doi: 10.1287/mnsc.2018.3253.
[49] F. Berzal, *Evaluation Metrics for Unsupervised Learning Algorithms*. Sebastopol, CA: O'Reilly Media, Inc., 2019.
[50] T. M. Charles, F. Calmon, and B. Ustun, "Predictive Multiplicity in Classification," *arXiv*, in *Proc. Int. Conf. Machine Learning*, pp. 6765–6774, 2019.