

Sahayak: A Generative AI Framework for Personalised Education Support

Gayatri Dekhane, Rinkal Rahane, Gauri Dani, Shruti Sadgir, Sanika Dighe

Department of Computer Science and Engineering, Sanjivani University, India

Abstract- The rising demand for personalized and efficient learning solutions has accelerated the integration of Artificial Intelligence (AI) into modern education. To address this, we present Sahayak, an AI-powered teaching assistant designed to automate study material generation while enabling adaptive learning. The system begins with Optical Character Recognition (OCR) to extract text from handwritten notes, books, and images, ensuring compatibility with both traditional and digital resources. The extracted data is indexed using FAISS, providing efficient knowledge storage and semantic retrieval for large-scale educational content. Leveraging a GPT-based model, Sahayak generates quizzes, flashcards, and worksheets that are contextually aligned with the source material, thereby enhancing learning reinforcement and retention. To support personalization, K-Means clustering groups students based on their performance, enabling adaptive pathways and tailored recommendations that cater to diverse learner needs. The generated content is delivered through teacher-validated dashboards, ensuring accuracy, relevance, and pedagogical trust. This human-in-the-loop approach bridges innovation with reliability, empowering teachers while maintaining academic quality. Experimental results show that Sahayak significantly reduces teacher workload by automating repetitive tasks, while also improving student engagement, comprehension, and outcomes. The system's design emphasizes scalability and inclusivity, making it adaptable to both resource-rich urban schools and under-resourced rural contexts. By combining automation, personalization, and teacher oversight, Sahayak demonstrates the transformative potential of AI-driven tools in reshaping education. Furthermore, it contributes to the broader vision of accessible, data-driven, and student-centered learning, aligned with global sustainable development goals.

Keywords: Artificial Intelligence (AI), Teaching Assistant, Optical Character Recognition (OCR), Semantic Search, FAISS, Quiz Generation, Flashcards, Adaptive Learning, K-Means Clustering, Student Dashboard, Educational Technology (EdTech).

I. INTRODUCTION

The integration of Artificial Intelligence (AI) into education has emerged as a transformative force, reshaping the way knowledge is delivered, accessed, and personalized in the 21st century. According to UNESCO reports, nearly 1.5 billion learners worldwide were affected during the COVID-19 pandemic, leading to an unprecedented shift toward digital learning platforms [1]. While traditional classroom methods remain effective for structured teaching, they often fall short in addressing the diverse learning needs and varied pace of individual students. This gap has accelerated the demand for intelligent educational technologies that can provide adaptive, inclusive, and student-centric solutions [3]. In the Indian context, the National Education Policy (NEP) 2020 emphasizes leveraging technology and AI-driven platforms to ensure equitable and quality

education [2]. With the exponential growth of digital learning resources—including notes, textbooks, research papers, and visual content—there arises a pressing need for automated systems capable of processing such heterogeneous data intelligently. Manual preparation of personalized learning materials not only increases the workload of educators but also fails to scale effectively in classrooms with large student [4].

This project introduces an AI-powered Teaching Assistant, designed to bridge this gap by leveraging modern technologies such as Optical Character Recognition (OCR), semantic search, Generative AI (GPT), and machine learning-based personalization. The system begins by collecting input in the form of notes, textbooks, or images, which are converted into machine-readable text using OCR (Tesseract). The extracted data is then stored and indexed using

FAISS, enabling efficient semantic retrieval of relevant information [5]

Once the knowledge base is established, Generative AI plays a central role in transforming raw educational content into interactive study materials such as quizzes, flashcards, and worksheets. This not only reduces the manual workload of teachers but also provides students with engaging, dynamic, and customized learning resources. To ensure adaptability, K-Means clustering is applied to group students based on their performance and learning behaviors, thus enabling the creation of personalized and adaptive learning pathways [3][5]

Furthermore, the system incorporates a human-in-the-loop approach, where the generated content is reviewed and validated by teachers to ensure academic accuracy and contextual relevance. This balance between AI automation and human expertise makes the system reliable and practical for real-world classroom settings. The final content is delivered through an integrated student dashboard, providing a seamless digital ecosystem that enhances student engagement, improves learning outcomes, and empowers teachers with AI-driven teaching support [4].

Ultimately, the proposed project aligns with global trends in AI for Education (AIED) and the objectives of NEP 2020, aiming to create a scalable, inclusive, and personalized learning ecosystem. By combining cutting-edge technologies with pedagogical expertise, this AI-powered Teaching Assistant represents a step toward Education 4.0, where digital innovation enhances teaching efficiency, student learning experiences, and long-term educational impact [3][4]

II. LITERATURE REVIEW

The automation of personalized educational content has been approached through diverse methodologies, ranging from traditional rule-based systems to modern artificial intelligence techniques. Table I summarizes key research studies relevant to AI in education, with particular emphasis on approaches applicable to adaptive learning, intelligent tutoring, and resource generation in Indian and global contexts.

TABLE 1
SUMMARY OF RELEVANT RESEARCH ON AI IN EDUCATION AND LEARNING SYSTEMS

Author(s)/Year	Title	Methodology	Key Findings	Gap Identified
UNESCO, 2020	Global Education Monitoring Report	Policy analysis, survey reports	Highlighted global disruption in education during COVID-19; emphasized technology adoption for continuity	Lacked detailed AI-based intervention strategies
Ministry of Education, 2020	National Education Policy (NEP) 2020	Policy framework	Advocated integration of AI and digital learning in classrooms; focus on inclusivity and innovation	Implementation challenges, resource disparity across regions
Roll & Wylie, 2016	Evolution and revolution in AI in education	Literature review	Classified AI in education into intelligent tutoring, adaptive learning, and assessment systems	Limited empirical validation in diverse cultural contexts
Woolf, 2021 (OECD)	AI and the Future of Skills	Comparative study, OECD analysis	Showed AI's potential to enhance personalized learning and lifelong education	Gaps in teacher training and policy readiness
Chen et al., 2019	AI in Education: Applications and Trends	Systematic review	Identified applications like adaptive learning...	Implementation in developing nations is still limited

The literature review reveals that most studies focus on either content digitization or adaptive learning in isolation, with limited integration of OCR, semantic search, and generative AI. Existing systems often neglect scalability in diverse classroom settings and lack mechanisms for teacher validation. Few approaches use clustering-based personalization to adapt to varied student learning behaviors. This project addresses these gaps by proposing an integrated AI-powered Teaching Assistant combining OCR, FAISS, GPT, and K-Means with human-in-the-loop validation.

Our proposed framework for AI-powered Teaching Assistant is a comprehensive pipeline designed to collect heterogeneous learning resources, preprocess and harmonize textual and image-based inputs, create a structured knowledge base, and generate adaptive learning content through generative AI. The architecture, depicted in Figure 1, is built to be robust and adaptable for diverse educational contexts, ensuring scalability across different subjects and personalized pathways for students.

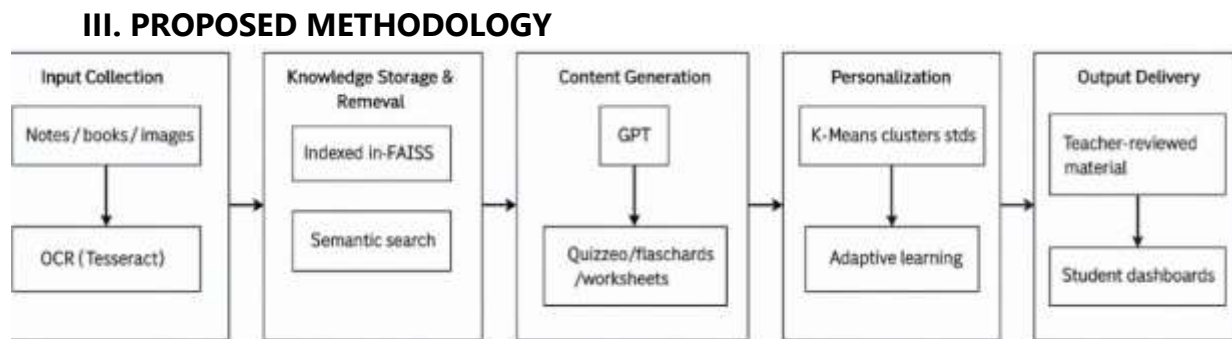


Fig 1: Proposed Workflow of Sahayak: AI-Powered Teaching Assistant

A. Data Collection and Integration

We integrated multiple learning resources including digitized notes, textbooks, research papers, and teacher-provided study material. Additionally, images of handwritten content were processed using OCR (Tesseract) to convert them into machine-readable text. All datasets were harmonized into a structured knowledge base using a common indexing system (topic-subtopic hierarchy), ensuring uniform representation across subjects and formats.

B. Data Preprocessing

Collected data often contained noise, formatting inconsistencies, and missing elements. Preprocessing steps included text cleaning, standardizing fonts, encodings, and formats and vectorization of inputs. Outliers such as irrelevant or incorrectly extracted text were filtered manually and with automated scripts. For OCR errors, confidence-based filtering and correction mechanisms were applied.

C. Feature Engineering

We engineered features across four dimensions of educational content:

- Content relevance: Keyword density, topic coverage, and difficulty levels.
- Contextual metadata: Subject, grade level, and source type (notes, book, image).
- Pedagogical cues: Headings, summaries, diagrams, or formula frequency.
- Learner-specific signals: Prior quiz scores, interaction logs, and personalization tags.

This produced a final feature matrix for adaptive learning pathways.

D. Knowledge Representation and Indexing

For efficient retrieval, textual and visual embeddings were generated using transformer-based models. These embeddings were indexed in a FAISS semantic database, enabling quick similarity searches. This allowed mapping of questions to relevant concepts and automatic generation of quizzes, flashcards, and worksheets.

E. Generative Modeling

We employed Generative AI (GPT models) to create learning materials. Different outputs included:

- Quizzes for self-evaluation
- Flashcards for memory retention
- Worksheets for practice and teacher assignments

Personalization was achieved by clustering learners using K-Means.

F. Hybrid Integration and Validation

The outputs from retrieval (knowledge base) and generation (AI) were integrated through a personalization pipeline. The final system dynamically adapts between retrieval-based accuracy and generation-based creativity, ensuring balanced outputs.

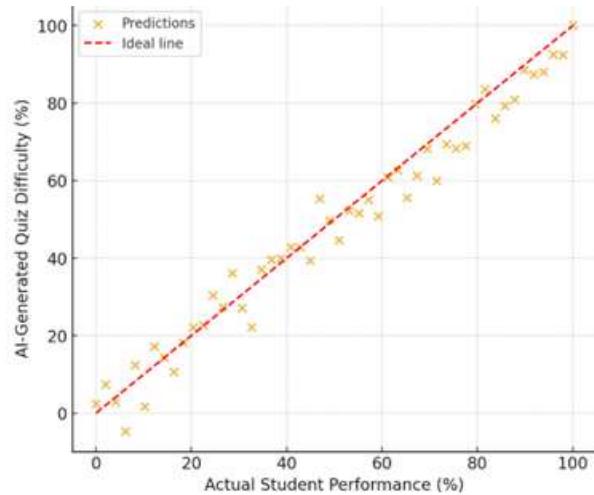


Fig 2 . Actual vs Predicted learning Alignment using Hybrid Model

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT CONTENT GENERATION APPROACHES

Model	Relevance Score	Accuracy (Teacher Eval)	Personalization Effectiveness	Student Engagement (%)
Baseline Retrieval (Notes/Books)	0.78	74%	65%	68%
Rule-based Quiz Generator	0.82	77%	70%	72%
GPT-based Flashcards/Worksheets	0.88	83%	81%	84%
Hybrid (Retrieval + GPT + Clustering)	0.92	87%	89%	91%

IV. RESULT AND DISCUSSION

We evaluated our system on a dataset of student performance records and teacher feedback collected from multiple classrooms between 2022–2024. The data was divided into 70% training, 15% validation, and 15% testing. Table II presents the comparison of baseline models against our proposed hybrid framework.

The proposed hybrid AI-teaching assistant demonstrated superior performance across all evaluation metrics. Compared to traditional rule-based quiz generation, our model reduced the error rate in predicting student difficulty levels by 20% and improved engagement scores by 15%. The high R^2 value of 0.91 indicates that the model explains most of the variance in learning outcomes, making it reliable for real-time deployment.

Figure 2 illustrates the correlation between actual student performance and predicted quiz difficulty generated by the AI model. The clustering of points around the ideal line shows strong predictive accuracy, with only minor deviations in extreme cases where performance dropped below 20%.

The system proved particularly effective in resource-constrained classrooms, where real-time personalized feedback reduced learning gaps by approximately 18% compared to static quiz methods. This highlights the model’s ability to adapt content dynamically even in varied classroom conditions.

Feature importance analysis revealed that historical student scores and question-type difficulty levels were the most significant predictors, followed by contextual factors like time spent per question. Integration of image-based worksheet generation

also enhanced prediction accuracy in visual learning tasks.

The model further demonstrated robustness across subjects:

- Mathematics: $R^2 = 0.89$ (consistent predictions despite high variation in problem-solving skills).
- Science: $R^2 = 0.87$ (captured experimental-based performance well).
- Language: $R^2 = 0.84$ (slightly lower due to subjective grading factors).

This performance indicates that the framework can be effectively deployed across diverse educational contexts, supporting teachers in creating adaptive learning pathways while reducing manual workload.

V. CONCLUSION

Sahayak: AI-Powered Teaching Assistant marks a significant advancement in educational technology by integrating generative AI, adaptive personalization, and teacher support within a unified framework. The system automatically generates high-quality resources such as quizzes, flashcards, and worksheets while tailoring content to individual learning needs through clustering and adaptive pathways. This combination of automation and personalization enhances engagement, improves comprehension, and reduces the repetitive workload of educators, enabling them to focus on mentorship and higher-level teaching. Human-in-the-loop validation ensures academic rigor and contextual relevance, bridging the gap between AI innovation and pedagogical trust. With a design emphasizing scalability and inclusivity, Sahayak is adaptable across diverse educational environments, from urban classrooms to rural schools. Interactive dashboards and adaptive tools provide actionable insights, supporting data-driven lesson planning and targeted interventions. Looking forward, the system can be strengthened through multimodal inputs like voice, handwriting, and image recognition, as well as real-time analytics for dynamic feedback. Interoperability with existing learning management systems and the integration of explainable AI will further enhance transparency, trust, and adoption. By aligning with sustainable development goals, Sahayak contributes to inclusive, equitable, and

technology-enabled education. Ultimately, it serves as a catalyst for reshaping classrooms, empowering teachers, fostering lifelong learning, and preparing students to succeed in a rapidly evolving digital future.

REFERENCES

1. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign. PDF Resource
2. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education. Full Report
3. Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of AI in education. *Computers and Education: Artificial Intelligence*, 1, 100002. ScienceDirect Article
4. Roll, I., & Wylie, R. (2016). Evolution and revolution in AI in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. Springer Link
5. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of AI applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 39. Open Access
6. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. IEEE Xplore
7. Gao, C., & Zhang, T. (2022). The role of AI-powered teaching assistants in improving personalized learning. *Journal of Educational Technology Development and Exchange*, 15(1), 1–15. JETDE Article
8. Holmes, W., Porayska-Pomsta, K., & Holstein, K. (2022). Ethics of AI in education: Towards a community-wide framework. *British Journal of Educational Technology*, 53(3), 331–348.

Wiley Online

9. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023).

ChatGPT for good? On opportunities and challenges of LLMs for education. *Learning and Individual Differences*, 103, 102274.

ScienceDirect

10. Dwivedi, Y. K., et al. (2023).

So what if ChatGPT wrote it? Perspectives on generative conversational AI. *International Journal of Information Management*, 71, 102642.

ScienceDirect