IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

A Personalized Autonomous AI Framework for Intelligent Study Planning, Material Summarization, and Adaptive Assessment

Mrs. Padmaja K[1], Himamshu B Bhat[2], Charanraj[3], Aniket Rathod[4]

East West Institute of Technology, Department of Information Science and Engineering Bengaluru, India[1-4]

Abstract: Personalized learning requires more than recommending content: it requires coordinating what a learner studies, when the learner studies it, how the material is explained, how understanding is checked, and how subsequent study is modified in response to performance. This paper presents a Personalized Autonomous AI Framework for Intelligent Study Planning, Material Summarization, and Adaptive Assessment. The implemented system is designed as a one-stop web-based learning environment in which a student registers or logs in, uploads multiple study materials such as PDF and Word documents, receives material-grounded summaries and a personalized study plan, interacts with a question-answering chatbot, completes milestone-based quizzes, receives weak-area identification, and obtains an automatically updated plan. A dashboard and progress-report layer records study and assessment information to support continuous self-regulated learning. The framework therefore connects content ingestion, educational summarization, grounded dialogue, planning, automatic assessment, learner-state estimation, and plan adaptation into a closed loop. The work is motivated by the observation that existing educational AI research frequently studies these capabilities as separate components. The paper reviews the relevant research described in the supplied literature, maps each research strand to the implemented modules, formalizes the end-to-end workflow, and discusses evaluation requirements for learning gain, summary quality, question quality, groundedness, plan quality, adaptation effectiveness, usability, and reproducibility. The manuscript also identifies research challenges including hallucination and omission in generated summaries, assessment calibration, sparse learner-interaction data, fairness, multimodal learning, long-term adaptation, and teacher oversight. The paper distinguishes the implemented system from claims that require further empirical validation and proposes a research agenda for evaluating whether closed-loop AI-supported study can improve learning outcomes and self-regulated learning compared with static study resources and non-adaptive planning.

I.INTRODUCTION

The rapid development of artificial intelligence has created new opportunities for educational systems to move beyond static repositories of learning resources. Traditional digital learning environments generally provide students with content, search facilities, assignments, and fixed schedules. These functions can be useful, but they often leave the learner responsible for deciding which material to study, how to divide limited time, how to verify understanding, and how to react when assessment reveals weaknesses. The central problem addressed in this work is therefore not simply the generation of educational content. It is the coordination of multiple learning-support functions around the changing state of an individual learner.

The proposed framework, titled “A Personalized Autonomous AI Framework for Intelligent Study Planning, Material Summarization, and Adaptive Assessment,” is designed as a one-stop environment for student learning. The student first accesses a web interface and creates or enters an account. After authentication, the student can upload study material, including PDF and Word documents. The system processes the material and makes it available as the basis for summaries, study planning, and question answering. Instead of requiring the student to manually transform a large collection of documents into a schedule, the framework generates a plan based on the available learning material and the intended study progression. The system also produces concise summaries that can be used as an entry point to more detailed study.

The learning process continues through an interactive question-answering interface. A student can ask questions about the uploaded material, and the chatbot is intended to answer using the learner’s document collection as the information source. After a milestone or defined portion of the plan is completed, the framework generates a quiz on the covered topics. The results of the assessment are analyzed at the topic level. Topics in which the student performs poorly are identified as weak areas, and the planner modifies subsequent study activities so that those areas receive additional attention. A dashboard and progress report provide a longitudinal view of completion, assessment performance, and learning progress.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

This closed-loop design is important because planning and assessment are normally treated as separate stages. A static study plan does not know whether the learner has mastered the material. A quiz that only produces a score does not necessarily change what the learner should study next. The framework connects the two: assessment becomes a source of evidence for replanning. In conceptual form, the system follows the cycle: material ingestion → summarization and planning → study and grounded dialogue → milestone assessment → weak-area identification → plan adaptation → subsequent study.

The objective of this paper is consequently to present the integrated framework as a research-oriented system rather than merely as a software application. The paper describes the problem, reviews the research strands underlying the architecture, presents the system design and workflow, explains the implementation, defines the learner adaptation loop, discusses evaluation methodology, and identifies research gaps and open questions. Where the supplied project document reports pilot findings, these are presented as reported prototype findings rather than as universally established evidence. A larger controlled evaluation is required before broad causal conclusions can be drawn.

A. Problem Statement

Students frequently encounter a combination of problems: large quantities of learning material, limited time, heterogeneous document formats, uncertainty about what to study first, difficulty extracting essential concepts, limited access to immediate explanations, and weak awareness of which concepts remain insufficiently understood. These problems are interconnected. A student who spends too much time on already-mastered topics may reach a deadline without covering weaker concepts. Conversely, a student who relies only on summaries may miss details that are necessary for deeper understanding. A useful learning assistant must therefore coordinate compression of information, planning, explanation, assessment, and adaptation.

The problem addressed by this research can be stated as follows: how can an AI-driven learning environment transform a learner's heterogeneous study materials and interaction history into a personalized, continuously updated study process in which planning, explanation, assessment, and remediation influence one another? The question is broader than automatic schedule generation because the schedule must react to evidence about learning.

B. Research Objectives

- Develop an integrated AI framework that connects study-material ingestion, summarization, planning, dialogue, assessment, weak-area detection, and plan adaptation.
- Provide a personalized study plan that organizes available material into study activities rather than presenting only a static document collection.
- Generate concise summaries that help students obtain an overview of uploaded learning material.
- Provide material-grounded question answering so that students can interactively clarify concepts from their own study corpus.
- Generate milestone-based assessments from covered learning content.
- Analyze assessment performance at the topic level and use the resulting weak-area information to update subsequent study activities.

C. Contributions

The primary contribution of this work is the design and implementation of an integrated learning loop. The framework brings together capabilities that are frequently investigated separately: automated planning, educational summarization, material-grounded dialogue, automatic question generation, learner-state analysis, adaptive scheduling, and progress visualization. The supplied reference manuscript explicitly frames this integration as the central novelty.

A second contribution is an explicit architecture in which assessment is not an endpoint. Assessment results are converted into topic-level evidence and fed back into the planner. A third contribution is the formulation of a practical evaluation methodology covering both educational outcomes and AI-system quality. A fourth contribution is a research-gap analysis that treats hallucination, assessment calibration, sparse learner data, fairness, multimodal learning, long-term adaptation, and teacher oversight as open areas rather than assuming that the current prototype has solved them.

II.RELATED WORK

A. AI-Powered Study Planning and Scheduling

AI-based study planning attempts to automate the organization of learning tasks according to constraints such as deadlines, available study time, topic difficulty, and learner preferences. The supplied literature describes Chavan et al. (2024) as a system that parses an uploaded syllabus PDF, extracts subjects, topics, deadlines, and estimated study times, and produces a tailored schedule. The reported system also includes progress tracking and adaptive suggestions. This

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

line of work is directly related to the planning module of the proposed framework because both approaches treat planning as a computational problem rather than a manually prepared calendar.
The reference manuscript also discusses Lingshetti et al. (2025), describing a smart study planner that uses behavioral information, machine learning, and educational-psychology considerations to adjust to time constraints and learning preferences. The cited work is presented as evidence that flexible schedules can be preferable to rigid schedules. The important distinction for the proposed framework is that the schedule is not intended to remain fixed after generation. It is connected to subsequent assessment evidence.

B. Automated Educational Material Summarization

Students often need to extract essential concepts from long notes, textbooks, lecture transcripts, or other documents. Automatic summarization can reduce the initial cognitive and time burden associated with reviewing large documents. The supplied literature discusses HumSum, attributed to Kolagar and Zarcone (2024), as a personalized lecture summarization tool based on GPT-4. The described approach uses a map-reduce strategy to process long transcripts and produce summaries influenced by explicit user preferences.
However, summarization introduces risks. A summary can omit important information, introduce unsupported statements, or simplify a concept to the point that its educational meaning changes. The supplied reference therefore identifies hallucination, factual accuracy, completeness, and educational appropriateness as continuing challenges.

C. Automatic Question Generation

Automatic Question Generation (AQG) is a well-established research direction concerned with generating questions from text, structured knowledge, or other source material. The supplied review describes transformer-based methods as increasingly important and notes that automatic question generation can support comprehension testing and formative assessment. The reference manuscript cites Le et al. (2017) and Lu and Lu (2021) as relevant reviews of the field.
The proposed framework generates quizzes after study milestones. This timing is important because assessment is linked to what the learner has recently studied. The quiz therefore serves both as an assessment mechanism and as a control signal for the adaptive planning process.

D. Intelligent Tutoring Systems and Knowledge Tracing

Intelligent Tutoring Systems (ITS) aim to adapt educational support to the learner. A central research problem is estimating what a learner currently knows. Knowledge tracing models infer latent knowledge states from sequences of interactions such as questions, answers, and feedback. The supplied manuscript describes Bayesian Knowledge Tracing as an early influential approach and Deep Knowledge Tracing as a neural extension capable of modeling sequential response patterns. More recent transformer-based approaches continue this direction.
Knowledge tracing is particularly relevant to the proposed framework because the system requires a mechanism for converting quiz performance into a learner state. In the simplest implementation, a topic can be assigned a performance score and classified as strong, moderate, or weak. A more research-oriented implementation can represent mastery as a probability that changes after each response. The planner can then allocate more study time to concepts with lower estimated mastery.

E. Educational Chatbots and Material-Grounded Question Answering

Conversational interfaces can change how students interact with learning resources. Rather than navigating documents manually, a learner can ask a natural-language question. The proposed framework uses a retrieval-augmented architecture in which uploaded materials are converted into searchable representations, relevant passages are retrieved, and a language model produces an answer using those passages as context. The supplied manuscript describes a vector-search pipeline using sentence embeddings and a vector store such as FAISS.
The central research concern is groundedness. A fluent answer is not necessarily a correct answer to the student's course material. For a study assistant, the answer should be traceable to the learner's material whenever the question concerns course-specific information. This motivates retrieval, citation of source passages, and evaluation of unsupported claims. The chatbot therefore functions as a bridge between static study material and interactive learning.

F. Learning Analytics and Progress Dashboards

Learning analytics focuses on collecting and interpreting information about learners, their interactions, and their progress. In the proposed framework, the dashboard is not merely a visualization layer. It provides a representation of learner state and makes progress visible to the student. The supplied manuscript identifies time spent, quiz accuracy, completed tasks, points, and milestone progress as examples of useful dashboard indicators.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

III. RESEARCH GAP AND MOTIVATION

The literature reviewed in the supplied reference demonstrates substantial progress in individual components of AI-supported learning. Study planners can generate schedules; LLMs can summarize educational content; automatic question-generation methods can produce assessment items; knowledge-tracing models can estimate learner knowledge; chatbots can provide conversational support; and dashboards can visualize progress. The central gap motivating the proposed framework is the limited integration of these capabilities into a single closed-loop learning process.

The integration gap should not be interpreted as evidence that no prior system has ever combined these features. Instead, it is a research question about how the combination should be designed and evaluated. Integration creates interactions that do not arise when modules are evaluated independently. For example, summary quality may influence quiz quality; quiz quality influences the accuracy of weak-area detection; weak-area detection influences replanning; and replanning influences the learner's subsequent behavior.

A second gap concerns the evaluation of adaptation. It is straightforward to state that a system changes a plan after a low score. It is more difficult to establish whether the change is pedagogically useful.

A third gap concerns groundedness and educational correctness. A language model can generate fluent summaries, explanations, and questions, but educational deployment requires evidence that these outputs are faithful to the source material and appropriate for the learner.

A fourth gap concerns sparse learner data. A learner may not produce enough interaction history for sophisticated knowledge tracing. The system therefore needs robust adaptation methods that work in both cold-start and data-rich settings.

A fifth gap concerns long-term personalization. A short pilot can demonstrate feasibility, but a genuinely autonomous study planner should respond to changing deadlines, study habits, course requirements, and mastery over longer periods.

A. Research Questions

• RQ1: Does an adaptive study plan that uses milestone assessment results improve learning gain compared with a static plan using the same study materials?
• RQ2: To what extent does material-grounded conversational support improve learners' ability to resolve conceptual questions during self-study?
• RQ3: How does summary quality influence subsequent quiz performance and learner confidence?
• RQ4: Can topic-level weak-area detection reliably identify concepts requiring additional study when learner interaction data are sparse?
• RQ5: What assessment frequency provides an effective balance between retrieval practice, study time, and learner burden?
• RQ6: How can automatically generated questions be calibrated for difficulty, coverage, correctness, and alignment with learning objectives?
• RQ7: How does the quality of the retrieved evidence affect the factual groundedness of chatbot responses?
• RQ8: Which learner signals are most useful for adaptive replanning: quiz accuracy, completion time, repeated errors, self-reported confidence, or combinations of these?

B. Testable Hypotheses

• H1: Students using an assessment-driven adaptive planner will demonstrate greater normalized learning gain than students using a static planner with equivalent source material.
• H2: Students receiving material-grounded conversational support will achieve higher performance on source-grounded conceptual questions than students using static document search alone.
• H3: Additional review sessions assigned to weak topics will improve subsequent performance on those topics relative to non-adaptive scheduling.
• H4: Generated assessments that satisfy source-faithfulness and learning-objective alignment criteria will provide stronger diagnostic value than assessments evaluated only by generic language-generation similarity.

IV. SYSTEM REQUIREMENTS AND DESIGN

A. Functional Requirements

• Registration and authentication: the student can create an account and securely access the learning environment.
• Material upload: the student can upload multiple study resources, including PDF and Word-based documents.
• Material processing: uploaded files are parsed into usable textual content and associated with the student's learning context.
• Summary generation: the system produces concise study-oriented summaries.
• Study planning: the system creates a schedule from available topics, study constraints, and learning progress.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

• Grounded chatbot: the student can ask questions about uploaded material.
• Performance analysis: assessment responses are associated with topics or concepts.
• Weak-area identification: low performance triggers a weak-topic state.
• Plan adaptation: subsequent study activities are modified to give weak areas additional attention.

B. Non-Functional Requirements

• Usability: the workflow should be understandable to students without technical knowledge.
• Reliability: document processing, plan generation, quiz generation, and progress recording should fail gracefully.
• Groundedness: material-based responses should preferentially use retrieved source evidence.
• Security: authentication and student learning data should be protected.
• Scalability: the architecture should support multiple documents and repeated interactions.
• Reproducibility: system configuration, prompts, model versions, datasets, and evaluation procedures should be recorded.
• Transparency: the system should distinguish generated recommendations from verified educational content.

V. PROPOSED FRAMEWORK ARCHITECTURE

The framework can be represented as seven connected layers: user and authentication, material ingestion, learning-content intelligence, personalized planning, conversational support, adaptive assessment, and analytics/adaptation. The architecture is intentionally closed-loop. The output of one module can become the input to another, and assessment results return to the planning layer.



Fig. 1. Conceptual architecture of the Personalized Autonomous AI Learning Framework.

A. User and Authentication Layer

The workflow begins with registration and login. Authentication establishes a persistent learner identity so that study materials, generated plans, quiz attempts, and progress information can be associated with the correct student. This layer is essential to personalization because adaptation requires continuity across sessions.
The user interface should expose a simple transition from authentication to material management. After login, the student should be able to upload documents, view existing learning material, open summaries, inspect the plan, access the chatbot, attempt assessments, and review progress.

B. Material Ingestion Layer

The ingestion layer accepts heterogeneous study material. The source manuscript describes PDF, Word, and plain-text input, with OCR or document parsing used to extract usable text.
After extraction, the system can normalize whitespace, remove irrelevant formatting, identify headings, segment chapters or topics, and associate text with document metadata. The result is a structured learning corpus. This corpus becomes the common source for summarization, planning, question answering, and assessment generation.

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.151190

A key research consideration is provenance. Each generated summary, answer, or question should ideally retain information about the source document and source section from which it was derived. Provenance makes debugging and educational review easier.

C. Content Summarization Layer

The summarization module converts long or detailed material into learner-oriented study representations. The supplied design proposes an LLM-based summarizer and describes a map-reduce strategy for long content: chunks are summarized and the intermediate summaries are synthesized into a final study guide.

The summary can be organized hierarchically: document summary, chapter summary, topic key points, definitions, important relationships, and revision points. Hierarchical summaries are particularly useful because a student may need a quick overview before deciding whether to read the full material.

The research version of the system should not evaluate summaries only by lexical overlap. Evaluation should consider factual consistency, completeness, relevance, readability, and usefulness for subsequent learning. This is important because a summary can use very different wording from an instructor reference while still being educationally accurate.

D. Personalized Planning Layer

The planner transforms the extracted topic structure and learner constraints into a schedule. The supplied manuscript describes factors such as available study time, topic difficulty, estimated time, and deadlines. Longer or harder topics may receive more study slots, while urgent deadlines can influence priority.

A simple planning formulation can be represented as a constrained allocation problem. Let $T = \{t1, t2, ..., tn\}$ be the set of topics, D the number of available study days, and H_d the available study hours on day d. The planner seeks an allocation $x_{t,d}$ representing study time for topic t on day d. Constraints can include $x_{t,d} \geq 0$, $\Sigma_t x_{t,d} \leq H_d$, and sufficient total allocation for each topic before its deadline. A priority function can combine deadline urgency, estimated difficulty, mastery, and previous exposure.

The adaptive extension changes this allocation after assessment. If mastery m_t decreases or falls below a threshold, the planner increases the allocation to topic t and may insert a revision session before progressing. Thus the schedule is a dynamic policy rather than a fixed calendar.

E. Material-Grounded Chatbot

The chatbot provides interactive learning support. The supplied framework describes embedding uploaded text into a vector store, retrieving relevant passages for a user query, and passing those passages to an LLM for answer generation. The intended interaction can be expressed as Query → Retrieval → Context Construction → Answer Generation → Optional Source Display. The retrieval stage reduces the probability that the model relies exclusively on its general pretrained knowledge. The answer should be considered grounded when its claims are supported by the retrieved material.

F. Milestone-Based Adaptive Assessment

The assessment module is activated after a milestone. A milestone can represent completion of a topic, chapter, group of chapters, or a planned study period. The generator uses covered material as the source for question construction.

The source manuscript describes a two-stage generation process: generation of question stems and correct answers followed by distractor generation for multiple-choice questions. It also suggests mixing question types and aligning questions with Bloom's taxonomy.

Assessment records should include question ID, topic, difficulty estimate, learner response, correctness, response time where available, and assessment attempt. These records provide the evidence needed for adaptation.

G. Weak-Point Detection and Knowledge Model

The weak-point detector maps assessment evidence to topic-level learning states. The simplest approach calculates accuracy per topic. A more advanced implementation maintains a mastery estimate that is updated after each response. The source manuscript proposes Bayesian Knowledge Tracing or a Deep Knowledge Tracing model for this purpose.

For a topic t, a simple diagnostic score can be defined as $A_t = C_t / N_t$, where C_t is the number of correct responses and N_t is the number of attempted questions associated with topic t. A weakness threshold θ can then be used to identify topics for remediation. However, such a threshold should not be treated as a universal truth. The number of questions, difficulty distribution, guessing probability, and confidence of the item labels all affect the reliability of A_t.

H. Adaptive Replanning Layer

The replanning layer is the central feedback mechanism. Once a weak topic is detected, the system modifies future study activities. The source manuscript describes inserting additional review sessions and reallocating time to weak topics.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

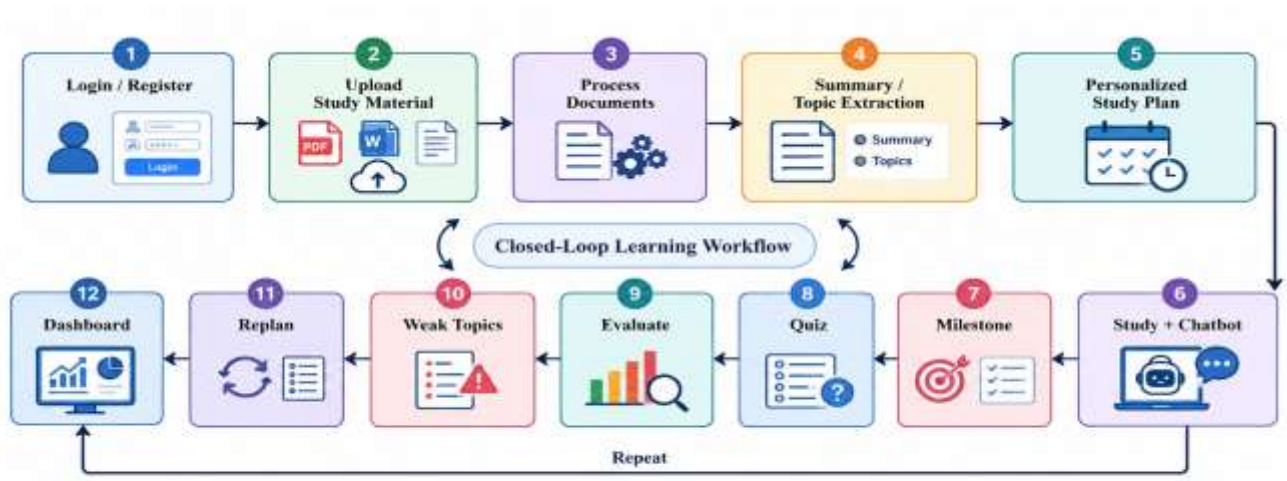


Fig. 2. Closed-loop learning workflow from material ingestion to adaptive replanning.

Adaptation can occur at several levels. At the topic level, the planner can increase time for a weak topic. At the sequence level, it can move a weak topic earlier. At the assessment level, it can schedule another quiz. At the explanation level, it can recommend the chatbot or a different summary. The research question is not whether the system can adapt, but which adaptation policy produces the best learning outcome for a given learner state.

I. Dashboard and Progress Reporting

The dashboard provides longitudinal visibility. The supplied manuscript identifies study hours, topic-level quiz accuracy, points, remaining tasks, and milestone progress as candidate metrics.

A research-oriented dashboard should distinguish activity metrics from learning metrics. Activity includes time spent, number of documents opened, and tasks completed. Learning includes mastery estimates, normalized

learning gain, retention, and performance on delayed tests. This distinction prevents the system from treating more interaction as automatically equivalent to better learning.

VI. END-TO-END WORKFLOW

The complete workflow can be described as a sequence of operational states. First, the learner authenticates. Second, the learner uploads study materials. Third, the system extracts and structures content. Fourth, summaries and topic representations are produced. Fifth, a study plan is generated. Sixth, the learner studies the assigned topics and can ask questions through the chatbot. Seventh, when a milestone is reached, the system generates and administers an assessment. Eighth, performance is mapped to topics. Ninth, weak areas are identified. Tenth, the planner updates the remaining schedule. Finally, the cycle repeats until the learner reaches the intended completion state.

A. Example Learning Scenario

Consider a student who uploads lecture notes for five modules of a subject and specifies a deadline and a limited number of hours available per day. The ingestion layer identifies the modules and topics. The summarization module produces an overview for each module.

The planner allocates study sessions across the
available days, considering the deadline and estimated workload.The student completes the first milestone and takes a quiz. Suppose the assessment contains questions from Modules 1 and 2. The student performs strongly on Module 1 but poorly on a subset of Module 2. The weak-area detector records Module 2 as requiring remediation. The planner then inserts a review session for the affected topics and may move the next assessment to occur after the additional review. The dashboard displays the change, allowing the learner to see why the plan has changed.

This example demonstrates the difference between static and adaptive planning. A static system would continue to the next scheduled topic regardless of the result. The proposed system uses assessment evidence as a control signal.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

VII. IMPLEMENTATION

The supplied project manuscript describes a web prototype with a React frontend and a Python backend, with MongoDB used for user and content data. It identifies Hugging Face Transformers, LangChain, PyMuPDF, and FAISS as key libraries, with GPT-4 used for summarization and answers, a fine-tuned T5 model for quiz generation, and a TensorFlow-based DKT implementation for knowledge tracing.

Because the implementation is modular, the system can replace individual AI components without changing the conceptual architecture. For example, the summarization model can be replaced by another language model, the vector store can be changed, or the learner model can be upgraded from threshold-based performance to a probabilistic knowledge-tracing model. This modularity is useful for controlled experiments because it enables ablation studies.

A. Data Model

A conceptual data model contains at least the following entities: User, Document, DocumentChunk, Topic, StudyPlan, StudySession, Summary, ChatInteraction, Quiz, Question, Attempt, TopicMastery, AdaptationEvent, and ProgressRecord. User identifies the learner. Document stores uploaded material. DocumentChunk stores retrievable text. Topic links content to a learning concept. StudyPlan contains scheduled activities. Quiz and Question represent assessments. Attempt records learner responses. TopicMastery stores the learner state. AdaptationEvent records why and how the plan changed.

The AdaptationEvent entity is especially useful for research. It can store the previous plan, triggering assessment evidence, identified weak topics, adaptation action, and resulting performance. This creates an audit trail for evaluating whether the autonomous loop behaves as intended.

B. Processing Pipeline

The document pipeline begins with file validation, followed by extraction and normalization. For PDFs, text can be extracted using a PDF parsing library. Word documents require document-level parsing. OCR can be used when a document contains scanned pages. Extracted text is then segmented for summarization and retrieval.

The summarization pipeline receives topic or document chunks, generates intermediate summaries, and combines them into a learner-facing representation. The chatbot pipeline uses a separate retrieval path so that questions can retrieve source passages rather than relying only on generated summaries.

C. Planner Logic

The planner can be implemented using heuristic scheduling, optimization, or a learning-based policy. A baseline heuristic can rank topics using a priority score $P_t = \alpha U_t + \beta D_t + \gamma(1-M_t) + \delta R_t$, where U_t represents urgency, D_t represents difficulty, M_t represents estimated mastery, and R_t represents remediation need. The coefficients can be tuned empirically. The exact function should not be assumed to be optimal; it is an experimental baseline.

After assessment, the remediation term R_t increases for weak topics. The planner then recomputes the remaining schedule subject to the learner's available time and deadline constraints. This creates a transparent adaptation mechanism that can be compared with more complex algorithms.

VIII. EVALUATION METHODOLOGY

An important distinction in evaluating this framework is between component-level quality and end-to-end learning effectiveness. A summary can be high quality even if the overall planner is ineffective. Conversely, a learner may improve despite imperfect summaries because repeated testing provides effective retrieval practice. Therefore, evaluation should include multiple layers.

A. Component-Level Evaluation

Summarization should be evaluated for factual consistency, completeness, relevance, readability, and usefulness. Automatic lexical metrics can be included where appropriate, but human evaluation should be used to determine whether important concepts are retained and whether unsupported claims are introduced.

Chatbot performance should be evaluated for retrieval quality, answer correctness, source groundedness, and abstention behavior. Test questions should include answerable questions, ambiguous questions, and questions whose answers are absent from the uploaded material. A robust system should avoid inventing source-based answers when evidence is unavailable.

Question generation should be evaluated for correctness, relevance to the source, uniqueness, difficulty, distractor quality, and learning-objective alignment. The source manuscript reports instructor review as one method for assessing generated questions.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

Planning should be evaluated for deadline feasibility, workload balance, topic coverage, adaptation responsiveness, and learner-rated realism. A generated plan that cannot be completed within available time should be considered infeasible even if it looks organized.

B. End-to-End Learning Evaluation

The most important experiment should compare at least two conditions: an adaptive AI condition and a non-adaptive or static condition. Both groups should receive the same core learning materials. The adaptive condition receives plan updates based on assessment. The control condition follows a fixed schedule or receives non-adaptive recommendations. A stronger design includes multiple conditions: static materials only; AI summaries and chatbot without adaptive replanning; adaptive planning without chatbot; and the full integrated system. This factorial or ablation design can help determine which components contribute to observed outcomes.

Learning gain can be measured using pre-test and post-test scores. A normalized gain can be computed as g = (Post - Pre)/(Maximum - Pre). A delayed retention test should be added when possible because immediate post-test improvement does not guarantee long-term retention.

C. Metrics

Dimension	Suggested Metric	Purpose
Learning	Pre/post gain	Measures change in knowledge
Retention	Delayed-test score	Measures persistence of learning
Adaptation	Weak-topic improvement	Measures remediation effectiveness
Planning	Plan feasibility	Checks deadline/time constraints
Planning	Plan adherence	Measures actual completion
Summary	Factual consistency	Checks unsupported information
Summary	Completeness	Checks retained key concepts
QA	Answer correctness	Measures factual accuracy
QA	Groundedness	Measures source support
Retrieval	Recall@k / precision@k	Measures evidence retrieval
AQG	Correctness	Checks answerability
AQG	Difficulty calibration	Checks intended challenge level
AQG	Distractor quality	Checks MCQ quality
Usability	SUS or equivalent	Measures usability
Engagement	Task completion / interaction rate	Measures usage
Efficiency	Latency and token/cost usage	Measures operational efficiency

TABLE I. PROPOSED EVALUATION METRICS FOR THE FRAMEWORK.

D. Statistical Analysis

For a controlled learner study, the analysis should report descriptive statistics, confidence intervals, effect sizes, and significance tests appropriate to the experimental design. If pre-test scores differ between groups, analysis of covariance or a mixed-effects model can be used to account for baseline performance. Repeated measures models are appropriate when learners complete multiple assessments over time. Effect sizes should be reported alongside p-values because statistical significance alone does not indicate practical educational importance.

E. Ablation Studies

Ablation is particularly important for an integrated framework. The following experiments can be conducted without rebuilding the entire system:

- Remove adaptive replanning while retaining summaries, chatbot, and quizzes.
- Remove the chatbot while retaining planning and assessment.
- Use a static summary instead of personalized summaries.
- Use random or fixed quiz questions instead of adaptive question selection.
- Use raw quiz accuracy instead of a learner-state model.
- Use a static plan and compare it with the adaptive plan.
- Compare source-grounded chatbot responses with unrestricted LLM responses.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

IX. REPORTED PROTOTYPE EVALUATION

The supplied reference manuscript reports a four-week pilot involving 30 undergraduate students and describes learning, usability, planning, summarization, question quality, adaptation, and chatbot observations. These values are reproduced here as reported in the supplied source document and should be independently verified before being presented as final publication results. The source reports a 15% average pre/post score improvement for the AI group compared with 8% for a static-material control, summary clarity of 4.3/5 and relevance of 4.1/5, approximately 0.65 ROUGE-L, 85% of generated questions rated good or okay, approximately 0.45 BLEU for AQG, 12% average improvement on weak topics following an additional review session, and 92% manually assessed chatbot factual correctness.

These observations are potentially useful as prototype evidence, but they should not be interpreted as definitive proof of causal educational benefit.

For the final thesis or journal submission, the authors should provide the participant allocation, inclusion criteria, course/domain, test construction, control condition, statistical tests, variance measures, confidence intervals, ethical/consent procedure, and exact experimental logs.

A. Interpretation of the Reported Findings

The reported learning-gain difference is consistent with the central hypothesis that combining planning and assessment may support learning, but the supplied information is not sufficient to establish which component caused the difference. The improvement could result from increased study time, repeated testing, greater engagement, the chatbot, planning quality, or a combination of these factors.

The reported summary ratings suggest that learners found the generated summaries useful, but subjective ratings do not establish factual completeness. Similarly, the reported question-quality percentage suggests that many questions were considered acceptable, but a stronger evaluation should analyze difficulty, distractor plausibility, topic coverage, and diagnostic value.

The reported weak-topic improvement is directly relevant to the adaptive-loop hypothesis. Nevertheless, future evaluation should include a matched non-adaptive comparison and delayed retention testing to determine whether the improvement is attributable to adaptive remediation and whether it persists.

X. RESEARCH GAPS AND OPEN PROBLEMS

A. Hallucination and Source Faithfulness

LLM-generated summaries and answers can introduce information that is not present in the source material. In an educational setting, this is particularly problematic because students may treat fluent output as authoritative. A research direction is to combine retrieval evidence, claim-level verification, confidence estimation, and abstention mechanisms.

B. Assessment Calibration

Automatic question generation does not automatically produce reliable assessment. Questions may be too easy, too difficult, ambiguous, repetitive, or poorly aligned with the intended concept. A research gap therefore exists in connecting generated-question quality with diagnostic usefulness.

C. Sparse Learner Data

A new learner may have only a few observations. A sophisticated neural learner model may therefore be unstable or poorly calibrated. Hybrid models combining prior knowledge, simple mastery estimates, item properties, and sequential interaction data may be more practical in cold-start settings.

D. Linking Summaries, Questions, and Plans

The system creates summaries, questions, and schedules from related material, but these outputs are often generated independently. A stronger framework could establish explicit traceability: each study task links to topics, each summary section links to source evidence, each quiz item links to a learning objective and source passage, and each adaptation decision links to assessment evidence.

E. Multimodal Educational Material

The current design is primarily text-oriented. Students frequently use diagrams, tables, mathematical expressions, slides, images, and video.

Future work can extend the ingestion layer to preserve visual structure and connect diagrams with textual explanations. However, multimodal expansion should be evaluated carefully because extraction errors can propagate into summaries and assessments.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

F. Fairness and Accessibility

Personalization may unintentionally favor learners who interact frequently, have stable internet access, or are comfortable with text-based interfaces. A learner who interacts less may be incorrectly classified as disengaged or low mastery.

Evaluation should examine whether adaptation decisions differ systematically across learner groups and whether the system provides equivalent learning opportunities under different interaction patterns.

XI. DISCUSSION

The framework illustrates a transition from component-level educational AI toward an integrated learning agent. The distinction is important. A summarizer produces information. A planner organizes activity. A chatbot answers questions. A quiz generator produces assessment. A learner model estimates knowledge. An adaptive planner changes future actions.

The closed-loop structure also introduces new failure modes. If the quiz generator produces a misleading question, the learner model can infer an incorrect weakness. The planner may then allocate time to a topic that the learner already understands. Similarly, if the chatbot provides an unsupported answer, the student may learn an incorrect concept, and a later assessment may reinforce the error. Therefore, reliability must be evaluated across the entire pipeline rather than at isolated components.

This motivates the concept of educational provenance. Every important AI-generated artifact should ideally be traceable: a summary to its source text, a chatbot answer to retrieved evidence, a question to its source and learning objective, a weak-point classification to the learner's responses, and a plan change to the detected learner state. Provenance can support debugging, teacher review, and research analysis.

Another important observation is that autonomy should not mean unrestricted automation. A responsible autonomous learning system should operate within explicit constraints. It should know when to abstain, when evidence is insufficient, and when human review is necessary. Autonomy is therefore better understood as automated decision-making under pedagogical and system constraints.

The framework also highlights the difference between personalization and adaptation. Personalization can mean tailoring content or presentation to preferences. Adaptation implies that the system changes because it has observed new evidence about the learner. The latter is central to the proposed architecture: quiz results alter the learner state, and the learner state alters the plan.

XII. LIMITATIONS

The first limitation is that the current research description does not establish a sufficiently large, diverse, independently replicated evaluation. The reported pilot in the supplied manuscript is preliminary and requires fuller methodological documentation before strong causal claims can be made.

The second limitation is dependence on language-model quality. Summaries, answers, and generated questions can inherit model errors. Retrieval can reduce unsupported generation but does not guarantee correctness.

The third limitation is the difficulty of estimating knowledge from a limited number of quiz interactions. Topic-level accuracy is useful as a simple diagnostic but is not equivalent to a complete cognitive model.

The fourth limitation is domain transfer. A framework that performs well on text-heavy subjects may not transfer directly to mathematics, programming, diagrams, laboratory procedures, or video-based courses.

The fifth limitation is the possibility that improved engagement is confused with improved learning. A student can spend more time interacting with a system without acquiring durable knowledge. Consequently, delayed assessments and objective learning measures are necessary.

The sixth limitation is that planning quality depends on accurate estimates of topic difficulty, workload, and available time. If these estimates are incorrect, an apparently intelligent schedule may still be impractical.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

Finally, privacy and responsible data management are essential because the system stores personal learning materials and performance history. Any deployment should apply appropriate data protection, access control, retention, and consent practices.

XIII. FUTURE WORK

Future work can proceed incrementally without replacing the existing framework. First, the evaluation can be strengthened through controlled experiments and ablation studies. Second, topic-level mastery can be upgraded from simple quiz accuracy to calibrated knowledge tracing. Third, question generation can be improved using explicit learning objectives and difficulty calibration. Fourth, chatbot responses can include source citations and abstention when retrieval evidence is insufficient.

A further extension is multi-agent orchestration. Separate specialized agents could manage document analysis, planning, assessment, learner modeling, and evaluation, while a coordinator maintains global constraints. However, additional agents should only be introduced when they provide measurable improvements; increased architectural complexity is not itself a research contribution.

Finally, teacher-in-the-loop controls can improve real-world adoption. Instructors can review generated assessment items, inspect weak-topic patterns at an aggregate level, and override unsuitable plan changes. This creates a human-AI collaboration model rather than a fully opaque autonomous tutor.

XIV. CONCLUSION

This paper presents a Personalized Autonomous AI Framework for Intelligent Study Planning, Material Summarization, and Adaptive Assessment. The framework addresses a practical educational problem: students must transform heterogeneous learning resources into an achievable study process while continuously determining what they understand and what requires further attention. The proposed system integrates material ingestion, summarization, personalized planning, material-grounded question answering, milestone-based quiz generation, weak-area detection, adaptive replanning, and progress visualization.

The supplied literature indicates that each individual component has significant prior research support, while the integration of these components raises additional research questions. The most important future work is consequently not simply adding more AI features, but demonstrating whether the integrated loop produces reliable and measurable learning benefits. This requires controlled comparisons, ablation studies, delayed retention testing, groundedness evaluation, assessment calibration, usability analysis, and transparent reporting.

The framework provides a foundation for such research. With rigorous empirical validation and explicit provenance linking source material, generated content, assessment evidence, learner state, and planning decisions, the system can serve not only as a practical study application but also as a platform for investigating autonomous and personalized learning processes.

ACKNOWLEDGMENT

The author acknowledges the academic guidance and resources used in the development of this research work. The supplied IJARCCE formatting template was used as the structural and formatting reference for this manuscript. The project reference document supplied for this work was used as the primary source for the framework description, reported prototype evaluation, research gaps, and cited literature.

REFERENCES

[1] M. P. S. Chavan, S. N. Nikam, D. P. Chikodi, M. R. D. Katkar and R. M. Mandhare, “AI-Based Study Planner,” International Journal of Creative Research Thoughts, vol. 12, no. 12, pp. d649–d653, Dec. 2024.
[2] G. Lingshetti, S. Pawar, M. Shinde and H. Sayyed, “AI-Powered Smart Study Planner: Enhancing Personalized Learning Through Intelligent Scheduling,” International Journal of Scientific Research in Science, Engineering and Technology, vol. 12, no. 3, pp. 148–155, May 2025, doi: 10.32628/IJSRSET2512320.
[3] Z. Kolagar and A. Zarcone, “HumSum: A Personalized Lecture Summarization Tool for Humanities Students Using LLMs,” in Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024), pp. 36–70, 2024, doi: 10.18653/v1/2024.personalize-1.4.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151190

[4] K. Keshamoni, "AI-Based Personalized Study Planner with Adaptive Reward Allocation for Enhancing Learning Effectiveness," SSRN preprint, 2026, doi: 10.2139/ssrn.6590378.
[5] N.-T. Le, T. Kojiri and N. Pinkwart, "Automatic Question Generation for Educational Applications – The State of Art," in Advanced Computational Methods for Knowledge Engineering, vol. 282, pp. 325–338, Springer, 2014, doi: 10.1007/978-3-319-06569-4_24.
[6] C.-Y. Lu and S.-E. Lu, "A Survey of Approaches to Automatic Question Generation: from 2019 to Early 2021," in Proceedings of the 33rd Conference on Computational Linguistics and Speech Processing (ROCLING 2021), pp. 151–162, 2021.
[7] X. Zhou, Z. Zhang, X. Xie et al., "Deep learning based knowledge tracing in intelligent tutoring systems," Scientific Reports, vol. 15, art. 21395, 2025, doi: 10.1038/s41598-025-07422-7.
[8] ASSISTments Data Mining Competition 2017, "Dataset," 2017.
[9] Riiid, "EdNet: A Large-Scale Hierarchical Dataset in Education," Riiid AIEd Challenge 2020 dataset.
[10] P. Rajpurkar et al., "SQuAD 2.0," Stanford Question Answering Dataset, 2018.
[11] OpenStax, "OpenStax: Free online textbooks and educational resources," Rice University.
[12] Wikipedia, "Wikipedia," Wikimedia Foundation.

APPENDIX A. LITERATURE-TO-SYSTEM MAPPING

Research Strand	Evidence in Supplied Literature	Framework Component	Research Need
Study planning	AI planners parse topics/deadlines and create schedules	Personalized Planner	Adaptive scheduling should be evaluated against static schedules
Summarization	Personalized LLM lecture summarization	Material Summarizer	Faithfulness, completeness, educational utility
AQG	Transformer and LLM-based question generation	Quiz Generator	Difficulty, distractor quality, diagnostic validity
Knowledge tracing	BKT/DKT and sequential response modeling	Weak-Point Detector	Sparse-data robustness and calibration
Educational chatbot	Retrieval + LLM responses	Material Chatbot	Groundedness and source attribution
Learning analytics	Progress tracking and adaptive reminders	Dashboard	Separate activity from learning outcomes
Adaptiv-e learning	Content and learning-path personalization	Replanner	Measure whether adaptation improves learning
Human oversight	Teacher review identified as future need	Governance layer	Human-AI collaboration

TABLE II. MAPPING OF RESEARCH STRANDS TO THE IMPLEMENTED FRAMEWORK.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151190

APPENDIX B. PROPOSED EXPERIMENTAL PROTOCOL

1) Recruit learners studying a common subject or course unit. 2) Collect a baseline pre-test. 3) Randomly assign learners to a static-planning condition and an adaptive-planning condition, or use a within-subject design when appropriate. 4) Provide equivalent source materials. 5) Allow the intervention period to run for a predefined duration. 6) Record study activity, chatbot interactions, quiz responses, plan changes, and completion. 7) Conduct immediate post-testing. 8) Conduct delayed retention testing. 9) Evaluate usability and perceived workload. 10) Analyze learning gain, retention, weak-topic improvement, plan adherence, and interaction metrics.

The protocol should predefine exclusion criteria, missing-data handling, test construction, randomization procedure, sample-size justification, and statistical analysis. If the study uses human participants, appropriate institutional approval and informed consent procedures should be followed.

APPENDIX C. PROPOSED FIGURES AND TABLES FOR THE FINAL PAPER

- Fig. 1. Conceptual architecture of the Personalized Autonomous AI Learning Framework.
- Fig. 2. Closed – loop learning workflow from material integration to adaptive replanning.
- TABLE I. PROPOSED EVALUATION METRICS FOR THE FRAMEWORK.
- TABLE II. MAPPING OF RESEARCH STRANDS TO THE IMPLEMENTED FRAMEWORK.

APPENDIX D. RESEARCH CONTRIBUTION STATEMENT

The research contribution can be stated as follows: “This work proposes and implements an integrated AI learning framework that closes the loop between educational material processing, personalized study planning, material-grounded conversational support, milestone-based automatic assessment, topic-level learner-state estimation, and adaptive replanning. Rather than introducing a new standalone summarization, question-generation, or knowledge-tracing model, the work investigates how these capabilities can be coordinated into an autonomous study process and proposes an evaluation methodology for measuring the educational value, reliability, groundedness, and adaptation effectiveness of the resulting system.”

 This work is licensed under a Creative Commons Attribution 4.0 International License