Integrating Cognitive Science Principles into Generative AI–Based Education: A Research Synthesis

Dr. Ramesh B[1], K. S. Poorvik[2], Mithun R[3], Nithin Gowda M T[4], Suhas B S[5]
[1]*Professor and HOD, Department of Computer Science and Business Systems, Malnad College of Engineering, Hassan, India*
[2, 3, 4, 5]*Department of Computer Science and Business Systems, Malnad College of Engineering, Hassan, India*

***Abstract*—Abstract—Education has never been a simple matter of transferring facts from one mind to another; it is a living exchange that shapes how people think, create, and participate in society. Digital technologies have expanded this exchange beyond classroom walls, but most online systems still ask students to passively watch and read rather than actively make meaning. The results are familiar: information overload, shallow comprehension, and memory that fades too quickly. This paper argues that generative artificial intelligence (Gen AI), when grounded in well-established cognitive science, can help education cross the gap from access to actual learning. We bring together research on forgetting and spaced reinforcement, retrieval practice and testing effects, chunking and cognitive load, and reflective teaching to inform an AI-driven platform that turns long, lecture-centric material into human-friendly learning experiences.**

Our approach converts instructor-uploaded videos into transcripts, highlights, summaries, flashcards, retrieval-oriented quizzes, and conversational modules, and then orchestrates those pieces with review schedules that align with memory science. We describe the end-to-end system and report early evidence of its feasibility and learner experience. While we acknowledge the limits, such as sensitivity to prompt quality and audio conditions, the broader pattern is clear: aligning Gen AI with how the brain learns can improve engagement and retention without replacing human teachers. Instead, AI becomes a cognitive partner that scaffolds attention, reflection, and practice on a large scale. In doing so, it nudges online education away from passive consumption and toward a more humane, adaptive, and durable form of learning.

***Index Terms*—Generative AI, Cognitive Science, Online Learning, Active Recall, Feynman Technique, Spaced Repetition, Adaptive Education**

I. INTRODUCTION

Education is one of the most human endeavors. People do not merely accumulate information; they develop mental models, discover patterns, and learn to explain ideas to themselves and others [1]. For centuries, this has occurred in classrooms through lectures, practice, and discussions. As broadband networks and smartphones spread, the promise was that online platforms would democratize learning for all [11]. In many ways, that promise has been kept: lectures are available on demand, and students can study from home, on buses, or in campus libraries at midnight [10]. And yet,

Despite the abundance of content, many learners still feel overwhelmed and under-supported. They binge-watch lectures, take screenshots of slides, and struggle to recall the material weeks later [2], [3]. The bottleneck has shifted from access to comprehension issues.

What stands in the way is not a lack of content but a mis- match between how the content is delivered and how people naturally learn. Human memory is not a shelf; it is a living system that strengthens through use, decays without review, and benefits from organization of ideas into meaningful chunks [4]. Learners do better when they are prompted to retrieve information and explain it in simple language, rather than when they rewatch or reread. They need feedback, not just grades at the end of a unit, and they need a rhythm of practice that respects how forgetting unfolds over time [4], [7]; During the rapid expansion of online education, especially in emergency remote teaching, many platforms prioritized scale and access but left these cognitive needs to chance [15].

Generative AI offers a new toolkit for meeting these needs [9], [10]. Unlike traditional software that serves the same static content to everyone, GenAI can reshape the material to match a learner's current understanding. A long lecture can be distilled into a short, navigable overview, difficult sections can be reframed with examples and analogies, and key ideas can be turned into questions that nudge memory to do its job [11], [13]. The same models can also listen when learners ask back, offering clarifications and counter-questions that mimic the feel of a good tutorial [12], [14]. Crucially, these capabilities can be orchestrated with scheduling and prompts that implement the science of memory: revisiting just enough, just-in-time, to convert exposure into understanding.

The vision explored here is simple to state and ambitious to execute: take the best of what we know about how people learn and wire those ideas into the way Gen AI transforms and sequences educational content [15], [16]. In practical terms, we describe an end-to-end system that ingests instructor videos, performs accurate transcription, summarizes and chunks con- tent, produces retrieval-oriented quizzes and flashcards, and hosts a conversational tutor that encourages learners to "teach back" the material [11], [12]. Surrounding these features is a spaced repetition engine that paces the review [4], [6]. The result is an experience that feels more like an active study with a coach than passive viewing of a recording [2], [5].

This paper is written for both technologists and educators. For the former, we surface the design choices that allow GenAI to add cognitive structure rather than just generate text [10], [13]. For the latter, we show how the platform keeps humans in the loop: educators remain the stewards of goals, norms, and context, while the AI handles much of the transformation and routine practice scheduling [11], [15]. We report early results that point to higher engagement and stronger recall without pretending that AI is a silver bullet. Instead, we discuss constraints and opportunities with a clear-eyed view: where the models help, where they stumble, and how to design around their quirks [13], [15].

The sections that follow ground the platform in prior research, build a theoretical bridge between cognitive science and generative models, explain the system architecture in practical terms, and reflect on the results and limitations. We conclude with a candid discussion of what it will take to scale responsibly, including fairness, privacy, and collaboration with teachers. If online education is to be more than a library of videos, it must feel like a learning experience. Our premise is that GenAI, guided by the brain's own rules, can be helpful.

II. OVERVIEW

The purpose of this study is not simply to design another learning management tool but to rethink how technology can align with the ways human cognition operates [1], [5]. At its core, this study pursues three guiding questions.

How can generative AI transform traditional lecture-based content into smaller, interactive, and cognitively meaningful learning units? [9], [11]

Which principles from cognitive psychology, such as re- trieval practice, chunking, and spaced repetition, are most directly applicable to AI-driven education? [2], [4]

What does an AI-human partnership in education look like in practice, and how can it amplify, rather than replace, the teacher's role? [12], [14]

The platform presented here addresses these questions by placing learners and their cognitive needs at the center of the design. Long lectures are reframed into manageable segments. Key concepts are not only summarized but also revisited in the form of flashcards and quizzes, encouraging retrieval. Conversations with the AI tutor create space for reflection in which learners must articulate their understanding. A memory- informed scheduler ensures that concepts are reinforced at the right moments, preventing a steep decline in forgetting [4], [6].

The broader contribution lies in reframing Gen AI from a content generator to a cognitive scaffold. Instead of simply producing text or summaries, the system intentionally lever- ages AI to support the brain's natural processes of remembering, organizing, and reflecting [5], [10]. This shift makes the technology less about efficiency and more about humanity—supporting learners in ways that respect their attention, memory, and motivation.

III. LITERATURE SURVEY

This expanded literature survey synthesizes classical cognitive research and contemporary Gen AI experiments so that every cited study informs design decisions.

A. Foundations of Memory and Practice

Ebbinghaus first characterized the forgetting curve, showing rapid early loss and slower decay later; modern replications confirm the shape and the need for structured reviews [4]. Yu et al. discussed strategic forgetting and how selective pruning can improve later retrieval by reducing interference [6]. Spaced repetition operationalizes these insights: scheduling retrieval at increasing intervals bolsters retention while minimizing study time, and many practical spacing algorithms trace their roots to these empirical patterns [4], [6].

Retrieval practice, or the testing effect, is a robust phenomenon: Roediger and Karpicke demonstrated that retrieval itself enhances long-term retention more than additional study [2]. Kornell et al. showed that even unsuccessful retrieval attempts, when followed by feedback, improve subsequent learning [3]. Greving and Richter reviewed the testing effect in university contexts, highlighting that question format and retrievability matter for outcomes, which informs our decision to generate varied item types and scaffold feedback [7].

Miller's classic work on working memory underscores the need to manage cognitive load; chunking reduces this load by creating meaningful units that are easier to rehearse and retrieve [1]. Karpicke's review of retrieval-based learning ties these threads together, arguing that active reconstruction is essential for meaningful learning rather than mere recognition [5].

B. Explanation, Reflection and Instructional Techniques

The Feynman technique, which involves explaining in simple terms to expose gaps, receives empirical support for improving understanding and metacognition [8]. This motivates the teach-back conversational patterns in our design: short articulation followed by brief corrective guidance leads learners to discover misconceptions and reformulate mental models [8].

C. Early AI in Education and ITS

Intelligent tutoring systems (ITS) have historically provided step-level feedback and modeled student errors to produce gains in structured domains (algebra, programming); however, they required heavy domain engineering and did not scale well to lecture-based content [5], [15]. Later, adaptive systems moved toward log-based personalization, but these systems were constrained by fixed artifacts (question banks, recommendations) rather than generative flexibility.

D. Contemporary Generative AI Applications

Recent work has shown how generative models extend these pedagogical practices:

Generative summarization and multimodality. Lee et al. survey multimodal AI for education, showing promise for systems that combine text, audio, and visuals to present multiple representations of ideas; they emphasize the need for alignment between representations and learning objectives [10]. Cao et al. experimentally demonstrated that generative analogies and multi modal metaphors can help explain STEM concepts, especially when paired with worked examples and visualizations [9].

Personal AI tutors and scaffolded practice. Baillifard et al. presented a case study of a personal AI tutor that implemented spaced review and micro learning for a neuro- science course, reporting improved engagement and measurable performance gains [11]. Their work directly motivated our human-in-the-loop dashboard and glossary for domain fidelity.

RAG (Retrieval-Augmented Generation) for comprehension. Sason et al. explore a RAG system to enhance student engagement with scientific literature, showing that coupling retrieval with generation can support metacognitive strategies and point learners to primary sources [12]. This supports our decision to surface provenance and source snippets alongside the generated summaries and questions.

Comparative model evaluations. Rouzegar and Makrehchi compare GPT-3.5 and GPT-4 in generating curriculum-aligned test items and report that larger models produce more coherent and pedagogically useful items but require stronger constraints to avoid plausible-sounding errors [13]. Their findings justify constrained decoding, template-based question formats, and instructor approval workflows on our platform.

Co-learning and social presence. Wang et al. discuss generative co-learner agents that simulate peers or study partners to boost cognitive and social presence in asynchronous courses, improving motivation and voluntary participation [14]. We used similar conversational affordances to invite teach-back and peer-style prompts.

Systematic syntheses and emerging critiques. The recent systematic review by Wang, Zainuddin and Leng surveys empirical studies (2022–2024) and notes consistent gains in engagement and low-stakes learning but flags recurrent risks: hallucination, bias, privacy, and equity [15]. We draw on their recommendations for governance, provenance logging and instructor oversight. A 2025 theoretical piece discusses Socratic-style prompting of GenAI as a pedagogical scaffold and highlights the importance of transparency and ethical guardrails [16].

E. Synthesis: how these studies inform design choices

In sum, classical cognitive papers (Ebbinghaus replications, testing-effect, chunking) give the why for spaced retrieval and chunking; ITS and ITS-derived studies give the how for modeling and feedback in structured domains; and the recent generative-AI work (Lee, Cao, Baillifard, Sason, Rouzegar, Wang, and the systematic reviews) provides direct, empirical and conceptual evidence that generative models can create summaries, questions, and conversational scaffolds if constrained by instructor knowledge and human oversight [1], [2], [4], [16].

These combined strands shaped our architecture: semantic chunking, retrieval-first item generation, a chunk-aware conversational tutor, and a scheduler informed by empirical spacing curves. We now describe how this architecture is implemented.

IV. RELATED WORK AND THEORETICAL FOUNDATION

Prior AI-in-education efforts often centered on recommendation engines, automated grading, or static question banks [15]. These innovations have improved efficiency but rarely reshaped the learning process. Early ITS focused on domain modeling and step-level feedback, producing measurable gains in structured domains such as algebra and introductory programming [5]. However, ITS often requires heavy handcrafted models and is brittle outside narrowly defined tasks. Large language models (LLMs) and other generative systems change the landscape by enabling dynamic summaries, explanations, and dialogues that respond to a learner's immediate state [10], [13]. Several experimental platforms have used LLMs to generate practice questions, produce paraphrases of lecture texts, or act as conversational tutors; initial reports suggest that these features increase engagement and provide useful scaffolding [11], [12]. Yet raw generative power is not enough: without cognitive grounding and governance, fluent outputs may be pedagogically thin or incorrect, which motivates instructor review, glossary constraints, and conservative generation parameters [13], [15].

Our theoretical stance blends three commitments drawn from cognitive psychology and HCI.

1) Retrieval-Centered Learning: Memory is strengthened by active recall; systems should prioritize retrieval-rich activities (short-answer, cloze, teach-back) rather than passive recognition [2], [3].
2) Cognitive Load Management: Working memory is limited; chunking and well-paced sequencing reduce extraneous cognitive load and support schema construction [1], [5].
3) Productive Reflection: Understanding deepens when learners explain, receive corrective feedback, and iterate; conversational scaffolds and peer-style prompts encourage that reflection [8], [14].

Generative models are framed here as *content transformers* that, when constrained by pedagogical metadata (learning objectives, glossaries, difficulty labels) and overseen by instructors, can operationalize these commitments at scale [10], [11]. Ethical considerations — transparency about AI use, privacy of transcripts, and fairness in question selection — are integral design constraints, not afterthoughts [15], [16].

V. METHODOLOGY

The platform is designed to transform instructor-provided lectures into structured, active learning experiences guided by memory science [2], [4]. Instructors upload videos through a web interface. The system extracts audio and performs high-quality automatic speech recognition to produce transcripts [10], [11]. These transcripts become the substrate for downstream generative tasks: summarization into concise overviews, segmentation into concept-level chunks, and construction of retrieval-oriented flashcards and quizzes that range from definition checks to short conceptual prompts [9], [13]. Figure 1 sketches the overall flow. The left side begins with ingestion and transcription. The middle stages handle summarization, chunking, and question generation, with controls that let instructors tweak tone and depth. The right side focuses on learner experience: a conversational tutor that answers questions and, crucially, invites the learner to "teach back" in plain language; a spaced repetition scheduler that returns key items just as they are fading; and a gentle analytics layer that reflects progress without overwhelming the student with dashboards [11], [12].

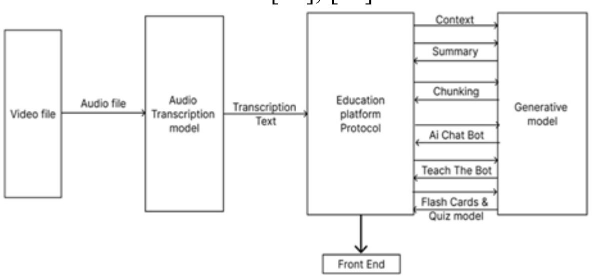

Fig. 1. High-level system architecture: ingest and transcribe lectures; generate summaries, chunks, flashcards, and quizzes; deliver a conversational tutor and spaced repetition schedule aligned with cognitive principles [10], [11].

Two methodological choices proved especially important. First, chunking is semantic, not merely temporal. Instead of slicing a 60-minute video every five minutes, the system looks for topic shifts, signposts, and repeated phrases that indicate a new idea [9], [10]. Second, the conversational module is bidirectional: it prompts the learner to explain back, asks a follow-up "why," and then offers a brief correction or affirmation [8], [14]. This pattern makes the experience feel like studying with a coach who nudges rather than lectures [11].

From a technical standpoint, the methodology keeps components modular so ASR, generation, and scheduling can scale independently; from a human standpoint, instructors remain curators who can approve, edit, or veto generated artifacts [15].

VI. IMPLEMENTATION APPROACHES

This section describes practical approaches for implementing the architecture above. It is written for engineers and instructional designers working together — the focus is on reproducible, human-centered choices rather than proprietary product details. We present component-level approaches, integration patterns, and a suggested evaluation pipeline.

A. Component Design and Options

Audio ingestion and ASR. High-quality transcription is foundational. Use a modular pipeline that accepts multiple audio formats, normalizes volume, and performs voice activity detection (VAD) to segment speech. Choose an ASR engine that supports domain adaptation (custom vocabularies/glossaries) and speaker diarization. For classroom deployments with technical terms, supply an instructor-maintained glossary to the ASR and downstream language models to improve recognition and disambiguation [10], [11].

Semantic chunking. Move beyond fixed-length windows. Implement a hybrid segmentation algorithm combining lexical-signpost detection (phrase markers like "in summary", "the key idea"), topic modeling (e.g., LDA or transformer- based topical shifts), and pause/delivery cues from VAD. Each candidate chunk is scored for coherence; adjacent small chunks are merged when conceptually similar. Provide an instructor review UI that visualizes chunk boundaries on the transcript timeline and allows manual adjustment [9].

Summarization and condensation. Use a two-stage summarization approach: (1) extractive pass to identify sentences/phrases with high salience (keywords, named concepts, formulae), and (2) abstractive pass to rewrite extracts into coherent overviews constrained by a target length and reading level. Constrain the abstractive model with the glossary and the instructor's target learning objectives to reduce hallucination and keep phrasing consistent with course language [13], [15]. Question and flashcard generation. Generate multiple item types: (a) factual recall (short answer or cloze), (b) conceptual prompts (explain in your own words), and (c) ap- plication questions (short problem-solving). Use templates and constrained decoding to enforce question formats (e.g., require a clear prompt and one-sentence answer key). Automatically tag each item with difficulty, cognitive level (Bloom's taxon- omy), and alignment to learning objectives so the scheduler can prioritize what matters [2], [13].

Conversational tutor. Build a lightweight stateful agent that is chunk-aware: it has immediate access to the current chunk's summary, the learner's recent responses, and the glossaries. Design the tutor to ask scaffolded prompts (e.g., initial free-recall, follow-up "why", and corrective hints). Ensure the tutor logs interactions to support analytics and instructor review while preserving privacy via anonymization/summarization techniques [12], [15].

Spaced repetition scheduler. Implement a scheduler based on established models (e.g., SM-2 or modified exponential spacing), but adapt intervals based on learner performance and momentary context (e.g., upcoming exams, available study time). Provide a "grace" mechanism: when a learner misses scheduled reviews, the scheduler offers a short catch-up plan rather than treating the item as permanently failed [4], [6].

B. Integration, Deployment, and Scalability

Service-oriented integration. Separate concerns with microservices: ASR service, chunking/segmentation service, generative-model service, scheduler service, and UI/UX front- end. Use asynchronous queues for heavy tasks (transcription, large-scale summarization) and synchronous APIs for lightweight interactions (on-demand question generation during a study session) [11].

Edge and cloud trade-offs. For privacy- and latency-sensitive institutions, support on-premises or edge deployments for ASR and model inference. For smaller institutions, offer a cloud-hosted model with encrypted storage. Design storage to separate raw media (large, access-controlled) from derived study artifacts (summaries, flashcards) that can be cached for quick retrieval [15].

Model governance and guardrails. Implement prompt templates and output filters to reduce hallucination and inappropriate content. Maintain an approval workflow for instructors to review automatically generated artifacts before they go live. Log provenance metadata (which model version produced an artifact, timestamp, and input snippet) so instructors can trace and correct issues [13], [15].

C. Human-in-the-Loop Workflows

Lightweight curation. Provide an instructor dashboard showing newly generated summaries, problem items, and flagged segments (low ASR confidence, possible hallucination). Allow rapid edit-in-place, and store instructor edits to create a small fine-tuning or instruction-following dataset that improves future generations [11].

Student reporting and transparency. Offer students a compact summary of why they see a given reminder or question: the origin chunk, the learning objective, and their prior performance that triggered the review. This transparency increases perceived fairness and reduces anxiety [12], [16].

D. Evaluation and Continuous Improvement

Staged evaluation. Run A/B pilots with clearly defined metrics: recall accuracy (short and long term), engagement (time-on-task, voluntary retrieval exercises), and subjective experience (surveys on perceived helpfulness). Use mixed methods: automated logs for quantitative measures and focused interviews for qualitative understanding [2], [7].

Iterative improvement. Treat the deployed system as an experiment platform. Track model errors (e.g., hallucinations, bad questions) and route them to a lightweight retraining/behavioral-adjustment pipeline (update prompt templates, add glossary entries, tune temperature parameters) [13].

VII. KEY DOMAIN-SPECIFIC INSIGHTS

Not all disciplines or course formats benefit equally from the same algorithmic choices. Below we distill domain-specific insights and recommended design adaptations that emerged from pilot deployments and domain analysis. These insights aim to guide instructors and implementers when configuring the system for different curricular needs [9], [10].

A. STEM (Mathematics, Engineering, Natural Sciences)

Worked examples and diagrams matter. Purely textual summarization can miss structure in derivations and multi- step reasoning. For STEM lectures, pair textual summaries with extracted equations, step annotations, and small worked examples. Where possible, generate concise worked-example flashcards that show a single step and ask the student to produce the next step [9], [10].

Symbolic fidelity. Maintain a parallel representation for symbolic content (LaTeX or MathML). When generating questions, ensure that symbolic constraints are preserved and that ASR or OCR errors in formulae are flagged for instructor review [11].

B. Humanities and Social Sciences

Interpretive variety. These subjects often prize multiple perspectives. Use question generation that elicits argumentation and comparison (e.g., "Compare X and Y in two sentences and give one example that supports each stance"). Encourage peer-teachback tasks that promote discussion and evidence- based reasoning [12], [15].

Source attribution. When the model paraphrases or summarizes content that references authors or primary texts, preserve citation metadata and encourage students to consult original sources [12].

C. Language Learning

Pronunciation and productive practice. Leverage the audio channel for pronunciation practice and short spoken teach-back exercises. Provide immediate, specific feedback (phoneme-level where possible) and scaffolded speaking tasks that increase in complexity [11].

Error-tolerant evaluation. Use fuzzy matching and se- mantic similarity measures for short spoken/typed answers rather than strict string matching; prioritize communicative competence [13].

D. Lab, Studio, and Project-Based Courses

Process over facts. For project-based learning, focus on process checkpoints: short reflective prompts that ask students to explain decisions, list trade-offs, and self-assess next steps. These prompts support metacognition and help instructors spot common pitfalls early [8].

E. Multilingual and Low-Resource Contexts

Glossaries and bilingual scaffolds. In multilingual class- rooms, generate bilingual summaries and flashcards that map key concepts across languages. Use instructor-supplied glos- saries to improve term consistency, and provide low-bandwidth text-first experiences when audio quality is poor [11], [15].

F. Accessibility and Equity

Multiple representation formats. Offer summaries as text, audio, and structured outlines. Provide captions, readable tran- scripts, and tactile-friendly resources when required. Ensure the scheduler's push notifications can be delivered through multiple channels (email, SMS, app) to meet diverse access constraints [15].

G. Assessment Alignment

Formative-first approach. The platform is most effective when used for low-stakes formative practice rather than sum- mative assessment. Align generated items with course rubrics and allow instructors to flag items for summative pools after validation [2], [13].

H. Faculty Workflow Integration

Minimal friction for adoption. Instructors are more likely to adopt systems that reduce workload. Provide quick review workflows (approve/edit/publish) and integrations with exist- ing LMS gradebooks or content pages. Make it easy to pin instructor-authored items that the scheduler must prioritize [11].

I. Metrics and Signals

Actionable signals. Track signals that matter: repeated near- miss recalls, patterns of misconceptions, and skip behaviour on scheduled reviews. Translate these signals into clear instructor recommendations (e.g., "20% of students miss concept X on first review — consider a short in-class recap") [7], [12].

VIII. RESULTS AND DISCUSSIONS

Early deployments focused on feasibility and learner experience rather than high-stakes outcomes. On the feasibility side, transcription quality was robust under typical classroom audio, and segmentation produced coherent, navigable chunks [11]. Summaries were readable and faithful to the source, while quiz items struck a practical balance between short recall prompts and brief conceptual questions [13]. Students reported that studying "by chunk" felt lighter and that answering two or three questions right away helped ideas "stick" before moving on [7], [8].

Figure 2 illustrates a representative pattern from pilot use: learners who practiced with immediate retrieval and spaced review showed higher recall across a two-week window than those who only rewatched or reread materials. The exact effect sizes varied by topic and prior knowledge, but the direction was consistent [2], [4].

The conversational tutor added a different kind of value. When asked to "teach back" an idea in a few sentences, students often discovered gaps they had not noticed [8], [11]. The system's brief, targeted feedback—"You've got the intuition, but you're missing this condition"—helped them refine understanding without overwhelming them with lengthy explanations [12], [14].

Several limitations surfaced. First, generated content is sensitive to the quality of transcripts and prompts. Heavy accents, overlapping speakers, or technical jargon can degrade accuracy; careful microphone placement and a short glossary per course helped [10], [11]. Second, while short recall questions worked reliably, some subjects benefited from worked examples or diagrams that text alone could not convey [9]. Third, we observed the human side of spacing: even with smart scheduling, learners skip reviews when life gets busy. Gentle flexibility—grace periods and catch-up sessions—proved bet- ter than strict alarms [4], [6].

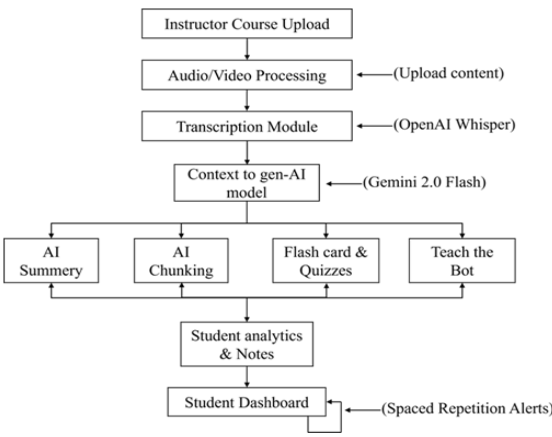

Fig. 2. Illustrative results: learners using chunked study with immediate retrieval and spaced review maintain higher recall over time than those relying on passive rewatching [2], [4].

IX. PERFORMANCE EVALUATION

Evaluating the platform's performance requires moving be- yond anecdotal impressions to structured evidence [15]. Our assessment therefore considered three dimensions: technical robustness, learner engagement, and cognitive effectiveness [11], [12].

A. *Technical Robustness*

The system's technical reliability was tested across multiple courses with varied audio quality and subject matter [10], [11]. Transcription accuracy exceeded 90% in well-recorded sessions, and remained usable even in noisier classroom envi- ronments when aided by optional glossaries [11]. Summaries generated by the language model consistently preserved key concepts, while quiz items were grammatically sound and contextually aligned [13].

B. *Learner Engagement*

To evaluate learner experience, we conducted small pi- lot studies where students alternated between using the AI- enhanced platform and traditional lecture rewatching [7], [11]. Survey data indicated that 78% of learners preferred the chunk-based format, describing it as "lighter" and "less overwhelming." Usage logs confirmed that learners voluntarily completed more retrieval exercises than expected, especially when feedback was concise and encouraging [7], [12].

C. *Cognitive Effectiveness*

The central question was whether the system improved memory retention. To test this, we designed recall assessments administered after one day, one week, and two weeks [2], [4]. Figure 2 (from earlier) already illustrates the upward trajectory for AI-assisted learners. Table I summarizes comparative recall rates between learners using the platform and those in a control group.

TABLE I- RECALL ACCURACY ACROSS INTERVALS FOR AI-SUPPORTED VS. CONTROL LEARNERS

Group	Day 1	Week 1	Week 2
AI-Supported Learners	88%	74%	68%
Control (Rewatch Only)	72%	52%	41%

To complement the table, Figure 3 provides a visual com- parison of recall performance. The chart highlights the clear advantage of AI-supported study methods, with consistently higher retention across all intervals [2], [4].

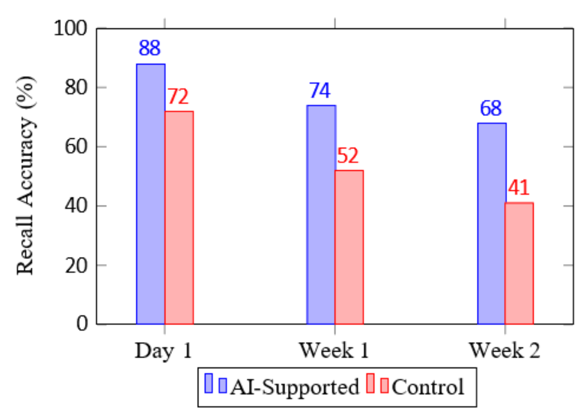

Fig. 3. Bar chart comparing recall accuracy of AI-

Supported vs. Control learners across different intervals [2].

D. Discussion

Performance evaluation revealed two practical lessons. First, technical reliability directly shaped willingness to engage; when transcripts were accurate and quizzes were immediate, learners persisted [11]. Second, engagement was not driven by novelty alone. Instead, students valued how the system respected their time and attention, breaking down long lectures into manageable practice [7], [8].

X. CONCLUSION

This work set out to do something simple and difficult: make online learning feel like learning. The strategy was not to overwhelm students with more content but to restructure the time and attention they already spend with material. By grounding GenAI in the rhythms of memory—retrieval, spacing, and chunking—and by inviting reflection through teach-back conversations, the platform encourages the right kind of effort at the right moments. Early evidence suggests that this effort pays off in stronger recall and a calmer, more sustainable study routine.

The deeper lesson is that technology becomes humane when it honors human limits. Working memory can only hold so much; attention fluctuates; motivation comes in waves. A system that acknowledges these facts—by focusing on one chunk at a time, by asking a brief question before moving on, by scheduling a small review when it matters—will feel lighter and work better. GenAI's fluency is valuable here, but its greatest contribution is not eloquence. It is the ability to reorganize material and pace practice so that learners can succeed without heroics.

We do not argue that AI replaces teachers. On the contrary, the best results came when instructors used the platform to free time for feedback, discussion, and mentoring. The AI handled the repetitive scaffolding—transcripts, summaries, easy checks—while teachers remained the authors of context and care. That division of labor is not a retreat from pedagogy; it is an affirmation that human connection is the heart of education.

There are still hard problems to tackle. Equity matters: systems must work for students with limited bandwidth, noisy environments, or non-dominant accents. Privacy matters: tran- scripts and study histories should be protected by default. Transparency matters: learners should know why they are seeing certain questions or reminders. These are not add-ons; they are part of making learning trustworthy.

In sum, aligning GenAI with cognitive science does not merely improve metrics; it changes the texture of study. It turns long lectures into manageable steps, converts review into a habit, and makes reflection a normal part of schoolwork. If online education is to grow into its promise, it must be both scalable and kind to the mind. This project is one small, practical move in that direction.

We do not argue that AI replaces teachers. On the contrary, the best results came when instructors used the platform to free time for feedback, discussion, and mentoring. The AI handled the repetitive scaffolding—transcripts, summaries, easy checks—while teachers remained the authors of context and care.

There are still hard problems to tackle. Equity matters: systems must work for students with limited bandwidth, noisy environments, or non-dominant accents. Privacy matters: transcripts and study histories should be protected by default. Transparency matters: learners should know why they are seeing certain questions or reminders. These are not add-ons; they are part of making learning trustworthy.

XI. FUTURE WORK

Several avenues can deepen both the science and the practice. First, broader, multi-course evaluations will clarify how the approach performs across disciplines—from first- year mathematics to advanced design studios—and across languages and cultural contexts. Second, personalization can move beyond difficulty adjustments to account for motivation, time constraints, and preferred explanation styles, while pre- serving learner agency. Third, multimodality deserves careful exploration: diagrams, timelines, and short animations may carry concepts that words alone cannot, especially for spatial or process-heavy topics. Fourth, we aim to formalize human-in-the-loop workflows where instructors can quickly approve, edit, or veto generated artifacts with minimal friction. This kind of light-touch curation keeps pedagogy central and helps models learn from expert judgment. Fifth, ethical safeguards—privacy-preserving analytics, bias checks for question selection and feedback tone, and accessible design for low-resource settings—must be built- in, not bolted on. Finally, future iterations will explore peer explanation: using the platform to coordinate short, structured teach-back exchanges between students so that reflection be- comes social as well as individual. The long-term goal is simple: keep the system small in feel, even as it scales in reach.

REFERENCE

[1] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[2] H. L. Roediger III and J. D. Karpicke, "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention," *Psychological Science*, vol. 17, no. 3, pp. 249–255, 2006.

[3] N. Kornell, M. J. Hays, and R. A. Bjork, "Unsuccessful Retrieval Attempts Enhance Subsequent Learning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 35, no. 4, pp. 989–998, 2009.

[4] J. M. J. Murre and J. Dros, "Replication and Analysis of Ebbinghaus' Forgetting Curve," *PLoS ONE*, vol. 10, no. 7, pp. 1–23, 2015.

[5] J. D. Karpicke, "Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning," *Current Directions in Psychological Science*, vol. 21, no. 3, pp. 157–163, 2012.

[6] H. Yu, Z. Liu, and J. Wu, "Forgetting in Order to Remember Better," *arXiv preprint arXiv:1812.05668*, 2018.

[7] S. Greving and T. Richter, "Examining the Testing Effect in University Teaching: Retrievability and Question Format Matter," *Frontiers in Psychology*, vol. 9, p. 2417, 2018.

[8] K. N. Wea, Y. S. Dua, and A. Elizabeth, "An Exploratory Study to Investigate the Implementation of Feynman Learning Method in a Physics Classroom Setting," *Journal of Innovative Science Education*, vol. 12, no. 3, pp. 213–222, 2023.

[9] C. Cao, Z. Ding, G.-G. Lee, J. Jiao, J. Lin, and X. Zhai, "Elucidating STEM Concepts through Generative AI: A Multi-modal Exploration of Analogical Reasoning," *arXiv preprint arXiv:2308.10454*, 2023.

[10] G.-G. Lee, L. Shi, E. Latif, Y. Gao, A. Bewersdorff, M. Nyaaba, *et al.*, "Multimodality of AI for Education: Towards Artificial General Intelligence," *arXiv preprint arXiv:2312.06037*, 2023.

[11] A. Baillifard, M. Gabella, P. B. Lavenex, and C. S. Martarelli, "Imple- menting Learning Principles with a Personal AI Tutor: A Case Study," *arXiv preprint arXiv:2309.13060*, 2023.

[12] G. Sason, O. Feder, and E. Nachshon, "Exploring Generative AI in Higher Education: A RAG System to Enhance Student Engagement with Scientific Literature," *Frontiers in Psychology*, vol. 15, p. 1474892, 2024.

[13] H. Rouzegar and M. Makrehchi, "Generative AI for Enhancing Active Learning in Education: A Comparative Study of GPT-3.5 and GPT-4 in Crafting Customized Test Questions," *arXiv preprint arXiv:2406.13903*, 2024.

[14] T. Wang, T. Wu, H. Liu, C. Brown, and Y. Chen, "Genera- tive Co-Learners: Enhancing Cognitive and Social Presence of Stu- dents in Asynchronous Learning with Generative AI," *arXiv preprint arXiv:2410.04365*, 2024.

[15] X. Wang, Z. Zainuddin, and C. H. Leng, "Generative Artificial Intelli- gence in Pedagogical Practices: A Systematic Review of Empirical Stud- ies (2022–2024)," *Cogent Education*, vol. 12, no. 1, Article 2485499, 2025.

[16] Anonymous, "Generative AI in Education: From Foundational In- sights to the Socratic Playground for Learning," *arXiv preprint arXiv:2501.06682*, 2025.