Study Pilot: An AI-Powered Platform for Personalized Learning through Retrieval-Augmented Generation on Diverse User Content

1st Ravi Teja Karnati
Dept. of Computer Science Engineering
Guru Nanak Institutions Technical Campus Hyderabad, India
ravitejakarnati5312@gmail.com

2nd Prof. (Dr.) Harish Kundra
Dean of Faculty and Student Affairs
Guru Nanak Institutions Technical Campus Hyderabad, India
drhkundra@gniindia.org

3rd P R S Santosh Naidu
Dept. of Electrical and Communication Engineering
Guru Nanak Institutions Technical Campus Hyderabad, India
prssn.raja@gmail.com

***Abstract*—**Personalized learning is a critical goal in modern education, yet existing digital platforms often struggle to adapt to the unique needs and materials of individual learners. Many AI-driven educational tools rely on static datasets, limiting their relevance to a user's specific curriculum or personal study resources. This paper presents Study Pilot, a comprehensive AI-powered platform designed to revolutionize personalized learning by enabling dynamic interaction with diverse content uploaded directly by the user, including PDFs, web links, and YouTube videos. At its core, Study Pilot employs Retrieval-Augmented Generation (RAG) [7], integrating advanced LLMs (via Ollama [1] and Groq API) with a vector database (ChromaDB) and other techniques to transform heterogeneous sources into co- hesive, interactive learning experiences. The platform supports multi-source course creation, contextual chat grounded in user content and the web, multiple language support via translation, collaborative course sharing, discovery of user-created resources, community forums, enhanced podcast generation for human- like audio [2], and a course rating system. Built with a Python Flask backend [3] and a JavaScript-based frontend [5], Study Pilot demonstrates a robust and scalable approach to making advanced, adaptable learning tools accessible to a wide range of users and content types. We present the system's architecture and methodology, and initial evaluations of the core document- based RAG functionality show promising results in accurate and relevant information retrieval for personalized learning.

Index Terms—Personalized Learning, Educational Technology, Retrieval-Augmented Generation, Large Language Models, AI in Education, Adaptive Learning, Multi-modal Learning, Collaborative Learning.

I. INTRODUCTION

The pursuit of personalized education, which tailors learning experiences to individual student needs, pace, and preferences, has been significantly advanced by the integration of artificial intelligence. However, a persistent challenge in the field of educational technology is the static nature of content in many AI-driven learning platforms. These systems often operate on pre-defined curricula or general knowledge bases, limiting their ability to engage with the specific textbooks, notes, or online resources that individual learners utilize in their studies. This disconnect hinders truly adaptive learning and restricts the potential for AI to provide deeply relevant support within the context of a student's actual learning materials.

The Study Pilot platform is developed as a response to this fundamental limitation. It is conceived as an AI-powered sys- tem that places the user's own diverse educational content at the center of the learning process. By leveraging sophisticated AI techniques, particularly Retrieval-Augmented Generation (RAG) [7], Study Pilot enables learners to transform static documents (PDFs), web pages (links), and even video content (YouTube) into interactive and personalized learning modules. The overarching objective of this project is to create a dynamic and adaptable learning environment that empowers users to learn from virtually any digital resource. This is achieved by building a system capable of processing, inte- grating, and generating educational interactions based directly on heterogeneous user-provided content. The platform aims to enhance accessibility and provide a more engaging and effective learning experience than traditional methods or AI
tools limited to fixed datasets.

The key contributions of Study Pilot, representing a significant advancement in personalized learning technology, include:

- A flexible system architecture designed to ingest and integrate diverse content modalities (PDFs, web links, YouTube) into a unified knowledge base for personalized learning.

DOI: 10.55041/IJSREM.NCFT007

- An advanced RAG implementation that combines document-based retrieval, contextual RAG techniques [6], and Web RAG [**?**] to provide highly relevant and context-aware responses grounded in both user-uploaded and external information.
- The development of a suite of AI-driven personalized learning features, including multi-source course creation, interactive chat, custom quiz generation, multi-language support, and enhanced podcast generation [2], all powered by the integrated RAG system [7].
- The incorporation of collaborative features such as course sharing, a find courses page for discovering user-created content, and a community forum, fostering a social learn- ing environment.
- An evaluation of the core document-based RAG functionality highlighting the feasibility and effectiveness of using user-uploaded content as a foundation for accurate and relevant AI-powered learning interactions.

Study Pilot has been developed through stages, beginning with a focus on PDF processing and core RAG chat functionality. The current iteration expands significantly on this foundation to incorporate a wider range of content types and collaborative features, realizing a more complete vision for a user-centric personalized learning platform.

II. Related Work

The integration of AI into educational platforms has been a subject of extensive research, with significant focus on personalized and adaptive learning systems. These systems aim to tailor content, pace, and feedback to individual student needs. Early approaches often relied on rule-based systems or knowledge graphs. The recent advancements in Large Lan- guage Models (LLMs) have shown great promise in creating interactive and conversational learning experiences. However, a common constraint for many existing AI tutors and learning assistants is their dependence on the data they were initially trained on or specific curated content libraries, limiting their ability to support learning from a user's unique collection of materials.

Retrieval-Augmented Generation (RAG) [7] has emerged as a powerful paradigm to overcome the limitations of LLMs based ledge sources to inform the generation process, leading to more accurate, factual, and contextually grounded responses. This approach helps to mitigate issues like hallucination and allows LLMs to provide more accurate and up-to-date infor- mation by grounding their generation in retrieved documents. Research in RAG has explored various aspects, including improving retrieval mechanisms through techniques like "Con- textual Retrieval" [6] and applying RAG to different domains. Study Pilot extends the application of RAG by focusing on the challenge of integrating arbitrary, unstructured PDF documents, web content (Web RAG) [**?**], and potentially extracting information from videos for a truly multi-source learning experience. The concept of Multimodal RAG is gaining traction, focusing on integrating information from various data formats like text, images, and audio. Study Pilot incorporates aspects of this by processing different content types and generating audio content (podcasts) [2].

Furthermore, the platform incorporates elements of collaborative learning, which has been shown to enhance student engagement and knowledge exchange. Existing collaborative learning platforms often leverage AI for features like content recommendations or facilitating discussions. Study Pilot combines individual personalized learning paths with features for collaborative course creation and community interaction, creating a holistic learning ecosystem. The potential role of AI agents in facilitating such environments is an area of ongoing research [8].

III. System Design and Architecture

Study Pilot is built upon a flexible and scalable layered architecture to accommodate diverse content types, AI processing workflows, and collaborative features. The system is organized into several interconnected layers: the Interface Layer, the Application/Process Layer, the Data Management Layer, and the Data Layer. While a detailed diagram is omitted for brevity, the conceptual architecture follows standard layered design principles.

A. Interface Layer

This layer is responsible for all user-facing interactions. It includes the web-based user interface built with HTML, CSS, and JavaScript, designed for intuitive navigation and presentation of diverse content and features. Tailwind CSS [5] is used for styling, providing a utility-first CSS framework for responsive design.

B. Application/Process Layer

This is the core of Study Pilot, housing the main processing logic and orchestrating interactions between different components. It includes modules for user authentication and manage- ment, multi-source content ingestion and processing, the RAG engine (incorporating document RAG and Web RAG [**?**]), personalized learning feature generation (chat, quizzes, pod- casts [2]), progress tracking, collaborative feature management (course sharing, community forum), and search/discovery of courses.

C. Data Management Layer

This layer acts as an intermediary between the application logic and the persistent data storage. It handles data validation, transformation, and formatting, ensuring data integrity and efficient access for the layers above.

D. Data Layer

This layer is responsible for the persistent storage of all data within the system. It utilizes **SQLite** for storing structured data such as user accounts, course metadata, chat histories, and progress tracking. A **ChromaDB** vector database is central to the RAG implementation, efficiently storing and enabling similarity search on the vector embeddings of all processed content.

IV. Methodology: Multi-Source Content Processing and RAG Implementation

A core innovation of Study Pilot lies in its ability to process and integrate diverse educational content beyond traditional static datasets. The methodology for handling multi-source content and powering AI interactions via RAG [7] is as follows:

A. *Content Ingestion*

Study Pilot supports the ingestion of content from multiple sources:

- **PDF Documents:** Utilizes libraries to extract text and potentially identify structural elements from uploaded PDF files.
- **Web Links:** Fetches content from provided URLs, focusing on extracting relevant text from web pages (Web RAG input) [**?**]. Integration with tools like Exa.ai [4] can enhance web content processing.

B. *Unified Content Representation*

Content from all sources is converted into a unified text-based representation where possible. For non-textual elements in PDFs or videos, descriptive text or summaries are generated or extracted.

C. *Content Segmentation and Chunking*

The unified text content is segmented into smaller chunks suitable for embedding and retrieval. Techniques like recursive character splitting are applied, with parameters adjusted based on the source type and content characteristics.

D. *Embedding Generation and Vectorization Optimization*

Text chunks are converted into high-dimensional vector embeddings using an embedding model (accessible via **Ollama** [1]). **Vectorization Optimization** techniques are applied to speed up the embedding process, particularly for large volumes of content. The 'nomic-embed-text' model is used for generating embeddings.

E. *Vector Database Indexing*

The generated vector embeddings, along with associated metadata (e.g., source type, original location, topic), are stored and indexed in the **ChromaDB** vector database. This enables efficient semantic search across all integrated content.

F. *RAG Implementation (Document RAG and Web RAG)*

Study Pilot employs an advanced RAG mechanism [7]:

- **Document RAG (Course RAG):** When a user interacts within a specific uploaded course, queries are embedded and used to retrieve the most relevant chunks *only* from that course's indexed content in ChromaDB. These retrieved chunks are then used as context for the LLM (via **Groq API**) to generate a grounded response. Contextual RAG techniques [6] are applied during chunking and retrieval to improve relevance.
- **Web RAG (Enhanced Rapid Learner):** For broader queries or in the "Rapid Learner" feature, the system can perform searches on the web to retrieve relevant information [**?**], potentially utilizing tools like Exa.ai [4]. This information is then used as context for the LLM, allowing the AI to draw upon both the user's content and external web sources for a more comprehensive response.

This multi-faceted approach to content processing and RAG allows Study Pilot to provide personalized learning experiences that are not limited by the format or origin of the user's educational materials.

V. Personalized Learning Features

Study Pilot leverages the RAG-powered core [7] to offer a suite of personalized learning features that enhance engagement and understanding of the user's uploaded content:

A. *Multi-Source Course Creation*

Users can create personalized courses by combining content from different sources – uploading PDF files, adding web links, and integrating information from YouTube videos. The system processes and integrates this content into a unified knowledge base for that specific course.

B. *Interactive Chat*

A core feature allowing users to engage in conversational interactions with the AI assistant about their course content. Leveraging Document RAG and potentially Web RAG [**?**], [4], the chat provides context-aware responses, explanations, and answers to questions based directly on the uploaded and integrated materials. The system can also suggest follow-up questions to guide the user's learning path.

C. *Custom Quiz Generation*

Based on the content of a selected topic within a user's created course, Study Pilot can automatically generate custom quizzes. The AI analyzes the relevant content to formu- late questions, providing personalized assessments for self- evaluation.

D. *Enhanced Podcast Generation*

Users can generate audio content (podcasts) based on summaries or explanations of topics from their course material. The system creates a script using the RAG engine and utilizes the **Play.ai API** [2] to convert the text into natural, human-like speech, offering an alternative learning modality.

E. *Multiple Language Support*

The platform includes features for translating content, enabling users to learn from materials in different languages or receive AI-generated content in their preferred language.

F. *Collaborative Course Sharing*

Users have the ability to share the courses they have created with other users on the platform, fostering a collaborative learning environment.

G. *Find Courses Page*

A discovery feature that allows users to explore courses created and shared by other users on the platform, expanding their access to educational resources.

H. *Community Forum*

Integrated forums where users can discuss specific courses or general topics, ask questions, and engage with peers and potentially instructors, enhancing the collaborative and community-driven aspects of the platform.

I. *Course Rating System*

Enables users to provide feedback on shared courses, helping others discover high-quality content and providing creators with valuable insights.

These features collectively create a dynamic and interactive learning experience that adapts to the user's content and learning preferences.

VI. Implementation Technologies

Study Pilot is implemented as a web application leveraging a modern technology stack to ensure performance, scalability, and a rich user experience.

The backend of the platform is built using **Python** with the **Flask** microframework [3], providing a flexible and efficient foundation for handling requests, managing data flows, and integrating external services.

Core AI functionalities are powered by integrating with advanced language models and embedding services. The **Groq API** is utilized for high-speed inference with LLMs (e.g., Llama3, Mistral), enabling rapid generation of chat responses, quiz questions, and podcast scripts. **Ollama** [1] is using models like 'nomic-embed-text', for converting text chunks into vector embeddings.

Data storage is handled by a combination of databases. **SQLite** is used for managing structured data such as user accounts, course metadata, chat histories, and progress track- ing. The **ChromaDB** vector database is central to the RAG implementation, efficiently storing and enabling similarity search on the vector embeddings of all processed content.

Frontend development utilizes standard web technologies: **HTML** for structure, **CSS** (with frameworks like Tailwind CSS [5]) for styling, and **JavaScript** for interactive elements and dynamic content updates.

Integration with external services is crucial for certain features. The **Play.ai API** [2] is used for converting generated text scripts into natural-sounding audio for the podcast feature. Libraries like PyPDF2 are used for PDF text extraction. Integration with web search tools like Exa.ai [4] supports the Web RAG functionality. LLMs from providers like Google Gemini [9] could also be integrated for various AI tasks.

VII. Evaluation and Results

Evaluation of the Study Pilot prototype focused on assessing the core functionalities, particularly the effectiveness of the document-based RAG implementation and the overall system performance. Due to the comprehensive scope of Study Pilot, initial evaluation concentrated on the foundational features inherited from the project's earlier stages.

A. *Performance Metrics*

Performance was measured through typical usage scenarios, focusing on responsiveness:

- Page Load Time: Assessed the time taken for key pages in the web interface to load.
- Content Processing Time (PDF): Measured the time required to ingest, process, and index a PDF document for RAG.
- AI Interaction Latency (Chat/Quiz Generation): Evaluated the time from user input (chat message or quiz request) to the generation of an AI response or quiz.

Observed performance metrics for the core document-based features were consistent with findings from the project's earlier stages, with chat responses typically generated within a few seconds and PDF processing time varying based on docu- ment length and complexity, averaging around 120 seconds for a moderate-sized file. Vectorization Optimization efforts contributed to reducing the time required for embedding generation during content ingestion.

B. *Accuracy and Relevance Assessment*

The accuracy and relevance of AI-generated content, particularly in the chat interface and quiz generation, were qualitatively assessed and through sample testing:

- Content Extraction Accuracy: Evaluated the correctness and completeness of text extracted from various PDF documents.
- RAG Response Relevance and Accuracy: Assessed if chat responses were factually correct and directly relevant to the user's query and the content of the uploaded document, demonstrating successful RAG grounding [7].
- Quiz Question Quality: Evaluated if generated quiz questions were pertinent to the specified topic and accurately reflected the content within the document.

Initial assessments of the core RAG functionality showed a high degree of relevance and accuracy in responses and generated quizzes when grounded in clear and well-structured PDF content. The RAG implementation successfully utilized the retrieved document chunks to inform the LLM's output, validating the approach for personalized content interaction.

Further, more extensive evaluation is planned to comprehensively assess the performance and effectiveness of the newly integrated multi-source processing, Web RAG [**?**], [4], collaborative features, and language support across a wider range of users and content types.

TABLE I
COMPARISON WITH ALTERNATIVE APPROACHES

Feature	Traditional LMS	Fixed AI Chatbot	Study Pilot
Content Source	Predefined	Training Data	User-Uploaded / Diverse
Personalization Depth	Limited	General	High (Content-Specific)
Content Flexibility	Low	Low	High
Interaction Type	Navigational/Ba	sic Conversational	Conversational + Assets
Knowledge Grounding	N/A	Internal Data	User’s Content + Web
Cost/Resource	Dev. Heavy	Train. Heavy	Processing/API

C. *Comparison with Alternative Approaches*

Compared to traditional LMS that rely on fixed content or simpler AI chatbots without external knowledge retrieval, Study Pilot offers distinct advantages, summarized in Table I: Study Pilot excels in content flexibility and personalized knowledge grounding, directly addressing a key limitation of existing methods by enabling interaction with any user-provided educational material.

VIII. DISCUSSION

The development of Study Pilot demonstrates the significant potential of leveraging RAG [7] and LLMs with diverse user- provided content to create a highly personalized and adaptable learning platform. The system’s architecture successfully inte- grates various content ingestion pipelines and utilizes a unified vector database to power AI features that are directly relevant to the user’s specific educational materials. The incorporation of Multi-Source Course Creation, Web RAG [**?**], [4], and Multi-language Support significantly expands the types of resources learners can utilize, moving beyond the limitations of platforms restricted to specific formats or languages.

The initial evaluation of the core document-based RAG functionality confirms the effectiveness of grounding LLMs in user-uploaded content for generating accurate and relevant learning interactions. This approach has the potential to rev- olutionize how individuals engage with their study materials, making learning more active, interactive, and tailored.

The inclusion of collaborative features such as course sharing, discovery, and community forums introduces a social dimension to the personalized learning experience. This acknowledges that learning often occurs in a social context and allows users to benefit from the resources and insights of others. The potential role of AI agents in facilitating such environments is an area of ongoing research [8].

However, integrating diverse content modalities presents ongoing challenges, particularly in accurately processing com- plex layouts in PDFs, extracting meaningful information from multimedia, and ensuring seamless indexing across disparate sources. The performance and scalability of processing ex- tremely large files or handling a very high volume of con- current users are areas that require continued optimization. While initial results for core RAG are promising, a thorough evaluation of the expanded features across varied content and user demographics is necessary to fully understand their impact and identify areas for improvement.

IX. CONCLUSION AND FUTURE WORK

Study Pilot represents a comprehensive and innovative approach to personalized education by enabling AI-powered learning directly from diverse user-uploaded content through the application of Retrieval-Augmented Generation [7]. By integrating PDFs, web links, and leveraging potential insights from videos, the platform overcomes key limitations of traditional e-learning systems and static AI tools. The suite of features, including multi-source course creation, contextual chat, custom quizzes, enhanced podcasts [2], multi-language support, and collaborative tools, creates a dynamic and user-centric learning environment.

The project successfully demonstrates the feasibility of building a platform that can adapt to virtually any educational resource a user provides, offering personalized interactions grounded in that specific content. Initial evaluation of the core RAG functionality validates the effectiveness of this approach in delivering accurate and relevant learning support.

Building upon the current capabilities of Study Pilot, future work will focus on:

- **Refining Multi-Source Processing:** Enhancing the accuracy and robustness of content extraction and processing pipelines for complex PDF layouts, multimedia content analysis, and dynamic web content.
- **Expanding AI Capabilities:** Exploring more advanced RAG techniques [7], including multi-modal RAG that can directly process and retrieve information across different modalities (text, images, audio) within a single interac- tion, and integrating more sophisticated adaptive learn- ing algorithms. **Deepening Collaboration:** Implementing advanced features for collaborative content creation, peer feedback mechanisms, and AI-moderated community discussions. The potential role of AI agents in these features could also be explored [8].
- **Comprehensive Evaluation:** Conducting extensive user studies and performance benchmarks across all features and content types to gather detailed data on effectiveness, usability, and scalability.
- **Streamlining Deployment and Accessibility:** Simpli-

fying the setup and deployment process for end-users and potentially exploring offline access capabilities for processed content.

Study Pilot, in its current form and with planned future enhancements, aims to establish a new standard for personalized digital education, making learning more accessible, engaging, and directly relevant to the individual learner’s journey and resources.

REFERENCES

[1] Ollama. https://github.com/ollama/ollama
[2] Play.ai documentation. https://docs.play.ai/documentation/get-started/o verview
[3] Flask documentation. https://flask.palletsprojects.com/en/stable/
[4] Exa.ai documentation. https://docs.exa.ai/reference/getting-started
[5] Tailwind documentation. https://v2.tailwindcss.com/docs
[6] Contextual RAG. https://www.anthropic.com/news/contextual-retrieval
[7] Retrieval-Augmented Generation for Large Language Models: A Survey. https://arxiv.org/abs/2312.10997
[8] AI Agents That Matter. https://arxiv.org/pdf/2407.01502
[9] Gemini Documentation. https://ai.google.dev/gemini-api/docs