LeanWise : An Autonomous Academic Assistant Using Agentic RAG

Mona Teja Kurakula

Volunteer Research Assistant

University of Southern California

mkurakul@usc.edu

Abstract

Students often face challenges in managing multiple courses, balancing dynamic deadlines, and extracting insights from unstructured materials such as PDFs, slides, notes, and spreadsheets. These difficulties hinder efficient study and increase the burden of planning and organization. Learn-Wise addresses these challenges through an agentic AI solution powered by Retrieval-Augmented Generation. The system processes diverse learning resources to generate personalized study plans, map key concepts across subjects, and dynamically update calendars and reminders as coursework evolves. By integrating planning, organization, and adaptive guidance, LearnWise provides timely support for reviewing notes, clarifying doubts, and tracking progress. Unlike static tools, the assistant continuously adapts to changing schedules, and individual needs, ensuring that students remain aligned with evolving academic needs.

1 Introduction

Artificial intelligence has rapidly transformed educational environments, offering new opportunities to personalize student support and streamline academic organization. Agentic AI systems that are capable of reasoning, planning, and adapting autonomously enable students to receive customized assistance across diverse subjects. Retrieval-Augmented Generation (RAG) methodologies further enhance the capabilities of Large Language Models (LLMs) by equipping them to understand, extract, and synthesize insights from unstructured academic materials, including lecture transcripts, notes, and syllabus. By combining agentic workflows and RAG, platforms like LearnWise aim to deliver highly adaptive academic assistants that support students with study planning, concept mapping, and timely guidance throughout the learning journey.This flexibility enables LearnWise not only to streamline how students engage with their studies but also to empower them to learn smarter and more efficiently.

1.1 Problem Statement

Even though digital course materials are widely available today, many students still struggle to keep them organized and easy to use. With so many types of content, it can be difficult for learners to find what they need and stay on top of their studies. These challenges are made more complex by differences in how courses are designed and by the fact that every student has unique learning needs as they are involved in different coursework respectively. As a result, getting timely study support or resolving questions often remains frustrating.

Current academic assistant tools do not fully address these issues. They are usually limited in how flexible they are and rely on fixed sets of information, which means they cannot adapt well to changing contexts or provide deeper, step-by-step support. To bridge this gap, there is a need for systems that make use of modern AI approaches, such as agentic retrieval-augmented generation (RAG) and large language models (LLMs). Such systems could offer more proactive, context-aware guidance that supports students in real time and adapts to their individual learning needs.

2 Related Work

Personalized academic assistance using AI has seen significant advances with the rise of agentic AI and retrieval-augmented generation (RAG) methods. Tools like HumSum Kolagar and Zarcone (2024) demonstrate how LLMs, combined with lecture transcripts and chain-of-thought reasoning, can produce tailored lecture summaries supporting student comprehension and study efficiency.

However, most existing approaches remain limited in scope. Many focus on a single modality. For example, working only with lecture transcripts, while offering little support for integrating study schedules or mapping concepts across different materials. Several large datasets, such as the MOOC Lecture Dataset Shahane (2025), Open Yale Courses ope (2024) (used in HumSum), and EduQuest Educational Data Mining Community (2024), provide valuable transcripts and open educational resources. Yet, comprehensive multi-format datasets that capture the full range of study workflows are still in the early stages of exploration and benchmarking.

3 Initial Thoughts and Progress

Initial progress is focused on comprehensive dataset exploration vital for the LearnWise Agent development. Key educational datasets have been identified and examined for their suitability in providing diverse, rich academic materials including transcripts, notes, and course metadata. Alongside, plans are underway to create and test custom non-confidential datasets tailored specifically for initial pipeline building to simulate real-world academic scenarios without privacy concerns.

In parallel, current Large Language Model (LLM) solutions such as open-source models like Llama-3 and frameworks like Ollama are being evaluated to assess their strengths and limitations in applying agentic reasoning and retrieval-augmented generation.

References

2024. Open yale courses. https://oyc.yale.edu/. Accessed: 2025-08-28.

Educational Data Mining Community. 2024. Eduquest: Lecture texts and questions for higher education. https://educationaldatamining.org/edm2024/proceedings/2024.EDM-posters.97/index.html. Accessed: 2025-08-28.

Zahra Kolagar and Alessandra Zarcone. 2024. Hum-Sum: A personalized lecture summarization tool for humanities students using LLMs. In *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, pages 36–70, St. Julians, Malta. Association for Computational Linguistics.

Saurabh Shahane. 2025. Mooc lecture dataset. https://www.kaggle.com/datasets/saurabhshahane/mooc-lecture-dataset. Accessed: 2025-08-28.