International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Integrating Generative AI into Eduportal: A Document-Grounded Academic Support Framework for College Erp Systems

Dhanush D. R.[1], Avish G. R.[2], Shrinkhala Kumari[3]
Department of Computer Applications
BMS College of Commerce and Management, Bengaluru, Karnataka, India
dhanushdrklr@gmail.com avishgr2006@gmail.com

Abstract - College ERP systems store academic records and administrative information, but they usually present it through fixed screens and menus, so users must still search across modules and documents to answer simple questions. This paper proposes an extension to EduPortal, a web-based college ERP, that adds a conversational assistant based on Generative AI (GenAI) and Retrieval-Augmented Generation (RAG). The assistant answers questions from students, teaching staff, and administrators using only approved institutional sources such as syllabi, academic calendars, policies, and learning materials. Access control is enforced at the retrieval stage so that the assistant cannot surface documents outside a user's role, and responses carry source references and abstain when evidence is insufficient. The paper contributes an architecture, a seven-step query workflow, design requirements for authorization, privacy, and human oversight, and an evaluation protocol covering correctness, groundedness, retrieval quality, abstention, authorization, latency, and user experience. The framework is presented as a design, and no implementation results are reported. Future work will implement the prototype and evaluate it with academic staff and students.

Keywords: Generative Artificial Intelligence, College ERP, EduPortal, Retrieval-Augmented Generation, Large Language Models, Academic Support

I. INTRODUCTION

Digital transformation has changed how higher education institutions store records, communicate with students, and coordinate academic activities. College ERP platforms can centralize functions such as student information, attendance, timetables, departments, subjects, announcements, fees, and learning resources.
Nevertheless, users may still need to search across different screens or documents to find an answer to a question about a course, timetable, institutional policy, or learning topic.

Generative Artificial Intelligence (GenAI), particularly large language models (LLMs), enables natural-language interaction and the generation of explanations, summaries, and responses. Research reviews identify opportunities for personalized learning and academic assistance while also highlighting concerns about factual reliability, privacy, bias, and academic integrity [1], [2]. These concerns are especially important when an AI assistant operates within a college information system.

This paper proposes a GenAI-enabled support framework for EduPortal. Rather than replacing the ERP database or existing role-based workflows, the proposed assistant acts as a conversational layer over authorized institutional knowledge. Retrieval-Augmented Generation is considered because it can provide relevant source material to a language model before it generates a response [3], [4].

The main contributions of this paper are: (1) a proposed architecture for connecting an AI assistant with a college ERP; (2) a RAG workflow with access-filtered retrieval over approved institutional documents; (3) design requirements for authorization, privacy, source attribution, and human oversight; and (4) an evaluation protocol,

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

with baselines, a reviewer procedure, and a pilot user study, for testing the prototype after implementation. The remainder of the paper reviews related work (Section II), states the problem and objectives (Section III), describes the proposed system (Section IV) and implementation considerations (Section V), presents the evaluation plan (Section VI), discusses privacy, security, and ethics (Section VII), and concludes (Section VIII).

II. LITERATURE REVIEW

A. Generative AI in Education

Systematic reviews report that GenAI is being explored for tutoring, explanation, content generation, feedback, and other academic tasks. These applications may support flexible learning, but their value depends on pedagogical design, verification, and responsible use [1], [2]. For EduPortal, this literature supports investigating an assistant that provides explanations and helps users locate official academic information, while avoiding unsupported claims that AI automatically improves learning outcomes.

B. Risks and Responsible Use

Responsible use of GenAI in education requires attention to accuracy and privacy. UNESCO recommends a human-centred approach to GenAI in education, including protection of data privacy, appropriate validation, and institutional policy development [5]. LLMs can produce plausible but inaccurate responses, a problem widely documented as hallucination [6], and information entered into AI services may raise privacy and confidentiality concerns. Retrieved documents are also an additional attack surface: instructions hidden in retrieved content can manipulate a model, an attack known as indirect prompt injection [7]. Accordingly, the proposed EduPortal assistant should not expose private student records to unauthorized users, should treat retrieved text as data rather than instructions, and should clearly communicate when it cannot answer reliably.

C. Retrieval-Augmented Generation

RAG combines information retrieval with text generation. Instead of relying only on the model's internal parameters, the system retrieves relevant passages from a designated knowledge source and provides them as context for response generation [3], [4]. A survey of RAG techniques describes how retrieval, augmentation, and generation components have been refined beyond the original formulation [8]. In an academic ERP, RAG could ground answers in approved syllabi, college rules, course materials, and other documents. Retrieval does not guarantee correctness; source quality, access filtering, relevance ranking, and answer verification remain necessary.

D. Research Gap and Motivation

The reviewed literature establishes the potential of GenAI in education and stresses reliability, privacy, and governance [1], [2], [5]. RAG-based educational chatbots, in turn, are typically described as question-answering tools over course or institutional content [3]. The design question addressed here is narrower and practical: how a conversational, document-grounded assistant can be added to an existing role-based college ERP without bypassing its authorization model, so that the permissions that govern the ERP also govern what the assistant can retrieve. The contribution of this paper is therefore a system design and evaluation protocol for this setting, rather than a claim of superiority over deployed institutional systems.

III. PROBLEM STATEMENT AND OBJECTIVES

Users of college ERP systems may encounter friction when locating institutional information across separate modules and documents. General-purpose chatbots, meanwhile, may not have access to the institution's current

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

and authorized information. A college-specific assistant must therefore combine conversational usability with controlled access to trustworthy sources.

The objectives of the proposed EduPortal framework are:

- Provide a natural-language interface for questions about approved academic and institutional information.
- Retrieve relevant passages from a maintained knowledge base before generating responses.
- Respect the existing roles and permissions of students, staff, and administrators.
- Provide source references where possible and abstain when evidence is insufficient.
- Define measurable tests for relevance, factual consistency, security, latency, and user experience.

IV. PROPOSED SYSTEM AND METHODOLOGY

A. Existing EduPortal Context

EduPortal is a web-based college ERP designed around role-based access and academic administration. Its technical stack comprises a web frontend, a Node.js/Express backend, and a PostgreSQL database hosted through Supabase. The AI layer described in this paper is a proposed extension and has not yet been implemented.

B. Proposed Architecture

The architecture, shown in Fig. 1, separates the ERP's transactional database from the AI knowledge-retrieval layer. In an offline stage, approved documents are collected, cleaned, divided into manageable passages, embedded, and indexed together with metadata describing their owner, version, and access scope. In the online stage, when a user asks a question, the backend first verifies the user's identity and role against the EduPortal database. The retrieval component then searches only passages whose access scope includes that role. The relevant passages are passed to the language model with instructions to answer from the supplied evidence, acknowledge uncertainty, and include source references. A validation step checks that the answer is supported by the cited passages before it is returned to the user.

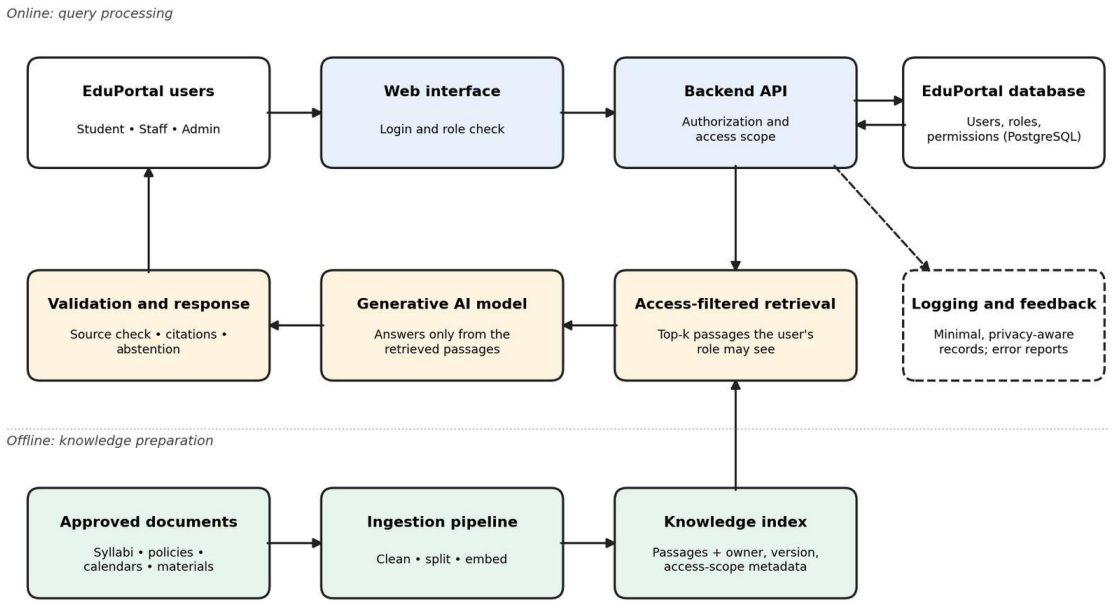


Fig. 1. Proposed architecture for a document-grounded Generative AI assistant in EduPortal.

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

C. Workflow

1. Authentication: The user signs in to EduPortal and is associated with a role and permitted data scope.
2. Question submission: The user asks a question using natural language.
3. Authorization and filtering: The backend checks the request and limits retrieval to approved sources.
4. Retrieval: The system searches the indexed syllabus, policies, timetable guidance, or learning resources for relevant passages.
5. Generation: The retrieved context and question are provided to the selected LLM to draft a response.
6. Validation and response: The system checks for source support, returns citations or document references where available, and indicates uncertainty when evidence is insufficient.
7. Logging and review: Minimal, privacy-aware logs may be retained for quality evaluation, subject to institutional policy.

D. Potential Use Cases

Student support: explain concepts from approved learning materials, locate syllabus information, and answer general questions about published academic rules.

Staff support: summarize approved course documents, locate teaching materials, and draft non-sensitive academic communications.

Administrative support: help authorized users find documented procedures and navigate ERP modules.

Scope and write actions: the framework in this paper is limited to questions answerable from approved documents. Questions about a user's own structured records, such as attendance, marks, or fee status, are outside its scope because they require tool-based database access under row-level security; this extension is left for future work. Actions that modify attendance, marks, fees, or student records should require explicit authorization and confirmation, and generated text alone should never change official records.

V. IMPLEMENTATION CONSIDERATIONS

A prototype can be implemented as a backend service integrated with the existing EduPortal API. The knowledge base should initially contain non-sensitive, institution-approved documents such as syllabi, academic calendars, course notes, and public policy documents. The ingestion pipeline extracts text, splits it into overlapping passages of a few hundred tokens, computes embeddings, and stores each passage with metadata such as document title, owner, version, department, semester, and access scope. Because EduPortal already uses PostgreSQL through Supabase, the passages and embeddings can be stored in the same database using the pgvector extension, which allows the access scope to be applied as a filter inside the retrieval query itself.

At query time, the system should apply authorization before or during retrieval, not merely hide restricted results in the user interface. Retrieval returns a small number of top-ranked passages (initially five). The prompt template instructs the model to (i) answer only from the supplied passages, (ii) cite the passage identifiers it used, (iii) distinguish sourced facts from general explanation, and (iv) state that the documents do not contain the answer when evidence is missing. Passage text is treated as data rather than as instructions, to reduce the risk of prompt injection [7]. The interface shows the sources used and provides a way for users to report incorrect responses.

Illustrative example (not an experimental result): a student asks for the minimum attendance required to appear for semester examinations. The backend verifies the student's role and retrieves passages from the published attendance policy whose access scope includes students. The model answers with the requirement stated in the

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

policy and cites the document and section. If the policy documents do not state a requirement, the assistant should say so and direct the student to the appropriate office instead of guessing.

The model provider, hosting approach, embedding model, and cost remain design choices to be fixed during implementation, based on institutional privacy requirements, response quality, latency, and available hardware. The main options are a hosted API and a locally deployed open-weight model; the latter avoids sending institutional text to an external service at the cost of hardware requirements.

VI. EVALUATION PLAN AND DISCUSSION

This paper proposes a framework, so no empirical results are reported. After implementation, the prototype will be evaluated using a curated set of questions whose expected answers and supporting sources have been reviewed by academic staff.

A. Suggested Evaluation Metrics

- Answer correctness: proportion of responses judged factually correct by reviewers.
- Groundedness: whether factual claims are supported by retrieved source passages. Reviewer judgment may be supplemented by automated RAG-evaluation tools such as RAGAs [9].
- Retrieval quality: whether the relevant source appears among the top retrieved passages.
- Abstention quality: whether the assistant declines or expresses uncertainty when the knowledge base lacks adequate evidence.
- Authorization tests: whether users are prevented from retrieving documents or data outside their permissions.
- Latency: time from question submission to completed answer.
- User experience: feedback from students and staff on clarity, relevance, and ease of use.

B. Test Design

The evaluation set should include straightforward questions, ambiguous questions, questions with outdated or conflicting documents, and questions for which no answer exists in the knowledge base. Security testing should include attempts to retrieve restricted information through direct questions and prompt-injection instructions contained in documents. Reviewers should record errors and classify them by retrieval failure, unsupported generation, stale source, authorization failure, or ambiguous wording.

Planned protocol: the test set will contain at least 100 questions distributed across the categories above and drawn from real syllabi and policy documents. The assistant will be compared with two baselines: keyword search over the same documents, and the same language model without retrieval. Two reviewers will independently label each response as correct, partially correct, incorrect, or appropriately abstained, and agreement will be reported using Cohen’s kappa. Authorization tests will use accounts for each role, and any retrieval of an out-of-scope document will count as a failure; the target is zero such failures. Usability will be assessed in a small pilot with students and staff using the System Usability Scale [10] together with open-ended comments.

C. Expected Benefits and Limitations

The expected benefit is a simpler conversational route to authorized institutional information, potentially reducing the effort of locating documents and understanding academic procedures; this remains to be measured. Limitations include incomplete or outdated source documents, retrieval errors, hallucinations [6], language-model costs, latency, and the need for staff oversight. The assistant should not be used to make consequential academic decisions without appropriate human review.

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

VII. PRIVACY, SECURITY, AND ETHICAL CONSIDERATIONS

The assistant should follow data-minimization principles and avoid sending personally identifiable student information to an external model unless the institution has approved the processing arrangement. Role-based access controls must be enforced on the server side. Sensitive records such as marks, attendance histories, fee details, and disciplinary information should not be indexed into a shared knowledge base without a documented authorization model. In India, the processing of students' personal data is additionally governed by the Digital Personal Data Protection Act, 2023 [11], and a deployment should be reviewed against it.

Institutional documents should have an owner, version, and review date so that outdated policies can be removed or superseded. Responses should link to the authoritative source where possible. Users should be told that AI-generated answers can contain errors, and official decisions must remain with authorized college personnel. To address academic-integrity concerns, the assistant should be configured to explain concepts and locate information rather than complete graded work, consistent with the institution's policy on acceptable AI use. These requirements align with human-centred guidance for GenAI in education [5].

VIII. CONCLUSION AND FUTURE SCOPE

This paper has proposed a framework for integrating Generative AI into EduPortal to provide conversational academic and institutional support. The approach combines a role-aware ERP interface, access-filtered retrieval from approved knowledge sources, language-model generation, source references, and validation controls. RAG offers a practical way to ground responses in institutional documents, but it does not eliminate hallucinations or replace the need for evaluation.

Future work will implement the prototype, build a controlled knowledge base, integrate authorization with existing EduPortal roles, and evaluate response quality and security using the protocol in Section VI. Further work may extend the assistant to structured, user-specific queries, such as attendance and timetable lookups, through tool-based access under row-level security, and may examine multilingual support (for example Kannada and Hindi), accessibility, personalized explanations, and staff-facing document assistance. Claims about efficiency or learning outcomes will be made only after suitable empirical evidence is collected.

REFERENCES

[1] A. Yusuf, N. Pervin, M. Román-González, and N. Md Noor, "Generative AI in education and research: A systematic mapping review," Review of Education, vol. 12, no. 2, e3489, 2024. doi: 10.1002/rev3.3489.
[2] L. Yan et al., "Practical and ethical challenges of large language models in education: A systematic scoping review," British Journal of Educational Technology, vol. 55, no. 1, pp. 90–112, 2024. doi: 10.1111/bjet.13370.
[3] J. Swacha and M. Gracel, "Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications," Applied Sciences, vol. 15, no. 8, art. 4234, 2025. doi: 10.3390/app15084234.
[4] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020. Available: https://arxiv.org/abs/2005.11401.
[5] F. Miao and W. Holmes, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023. doi: 10.54675/EWZM9535.
[6] Z. Ji et al., "Survey of hallucination in natural language generation," ACM Computing Surveys, vol. 55, no. 12, art. 248, 2023. doi: 10.1145/3571730.
[7] K. Greshake, S. Abdelnabi, S. Mishra, C. Endres, T. Holz, and M. Fritz, "Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection," in Proc. 16th ACM Workshop on Artificial Intelligence and Security (AISec), 2023. Available: https://arxiv.org/abs/2302.12173.
[8] Y. Gao et al., "Retrieval-augmented generation for large language models: A survey," arXiv:2312.10997, 2023. Available: https://arxiv.org/abs/2312.10997.
[9] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert, "RAGAs: Automated evaluation of retrieval augmented generation," arXiv:2309.15217, 2023. Available: https://arxiv.org/abs/2309.15217.
[10] J. Brooke, "SUS: A quick and dirty usability scale," in Usability Evaluation in Industry, P. W. Jordan et al., Eds. London, U.K.: Taylor & Francis, 1996, pp. 189–194.
[11] Government of India, The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023), New Delhi, India, 2023.

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