**Notionary - An AI-Powered Interactive Learning Assistant for Classrooms**

**Abstract**

Notionary is an AI-powered platform designed to revolutionize how students interact with learning material by enabling contextual question-answering from uploaded classroom documents. It combines local embedding generation using Ollama nomic-embed-text model with powerful LLMs like Google Gemini to provide highly relevant, strictly document-bound answers. With advanced document retrieval, robust user isolation through Supabase Row-Level Security and a clean frontend interface, Notionary presents a scalable and privacy-focused solution to enhance student engagement and clarity in digital learning environments. By offering localized AI capabilities, Notionary also mitigates risks related to privacy, latency and data ownership that plague traditional cloud-based EdTech tools.

**1. Introduction**

In traditional educational systems, students often struggle to get timely and accurate help with study materials outside the classroom. Even AI tools like ChatGPT or Gemini often hallucinate or provide generic responses unless tightly controlled. Notionary was built to address this gap by ensuring that AI responds *only* based on specific class material uploaded by the student or educator, providing accurate, reliable and document-specific help.

The core problem is twofold:

* Lack of personalized and document-specific query resolution
* Potential privacy leaks and hallucinations from generic AI models.

Notionary tackles both using a novel combination of secure vector search, strict LLM prompting and modular backend services. Additionally, Notionary integrates seamlessly into existing educational ecosystems and can scale from small classroom deployments to institution-wide installations.

**2. Methodology**

Notionary employs a combination of local and cloud technologies for an optimized educational assistant:

* **Document Upload & Processing**: PDFs are uploaded via a frontend UI. Text is extracted and intelligently segmented into chunks that preserve context.
* **Embeddings Generation**: Using Ollama's nomic-embed-text, 768-dimension vector embeddings are computed locally, ensuring performance and student data privacy.
* **Vector Storage & Retrieval**: Supabase's PostgreSQL with pgvector extension stores embeddings. Vector similarity search identifies the most relevant content chunks.
* **Language Model Integration**: Google Gemini 1.5 Flash is used to generate precise, context-aware answers. Prompt engineering ensures responses are grounded in retrieved content.
* **Strict Prompt Enforcement**: When document-specific mode is active and no context is found, the LLM is instructed to refrain from generating any answer beyond the scope of the document.
* **Security & Isolation**: Supabase’s Row-Level Security policies ensure users only access their own content. JWT-based authentication provides additional integrity.

**3. Optimization Techniques**

To enhance system efficiency and reliability, Notionary incorporates multiple optimization layers:

* **High-Dimensional Embeddings**: 768-dimensional vectors offer better semantic richness than typical 384-dimension setups.
* **Chunking Strategy**: Document chunks are generated with overlap and coherence awareness to maintain meaning across splits.
* **Relevant Chunk Filtering**: Only chunks with similarity > 0.4 are passed to the LLM, reducing noise and token consumption.
* **Service Role Bypass for Verified Users**: Certain DB queries use elevated permissions *only* after verifying document ownership, balancing security and accessibility.
* **API Response Tuning**: Responses are structured with quick answer, detailed explanation, examples and step-by-step reasoning, improving learning outcomes.

**4. Results & Discussion**

Notionary has been evaluated across academic subjects including math, science, and history, using real-world study notes and scanned exam papers.

**Results:**

* 100% user isolation verified through RLS enforcement
* Accurate source-linked answers retrieved from user PDFs
* No hallucinations detected in strict document mode
* PDF processing latency reduced by 60% due to local embeddings
* High user satisfaction in test groups

**Discussion:**  
Notionary’s emphasis on document fidelity significantly reduces the cognitive burden on students, allowing them to focus on clarification rather than discovery. Teachers can upload curated materials and know that students will receive only relevant, bounded responses. The RLS model also enables this system to be deployed in multi-user environments such as classrooms or institutions without compromising data integrity.

Use cases extend beyond classrooms:

* Coaching centres
* Universities with remote learning platforms
* Internal training document QA for corporate teams

Limitations include lack of support for real-time collaborative annotation and current dependence on PDF format. Future work will involve extending support to DOCX, HTML along with voice-driven interfaces and mobile apps.

**5. Conclusion**

Notionary provides a highly effective, document-bound and secure way for learners to engage with educational material using the power of AI. It bridges the gap between static documents and dynamic learning by combining trusted vector search, strict access control and intelligent language generation. Its modular, scalable architecture ensures easy extensibility, making it not only a solution for today’s classrooms but a platform for the future of interactive education.

**Created by,**

Aravind Chintalapati (RA2311003010479)

Nikunj Mathur (RA2311003010468)