IMPLEMENTATION: The system is implemented through a multi-stage workflow involving document processing, embedding generation, retrieval, and output synthesis. Uploaded educational materials are first extracted and cleaned to produce raw textual content. This content is segmented into smaller, semantically meaningful units and converted into dense embeddings using sentence-transformer models. These embeddings are stored in a Chroma DB vector database for efficient retrieval. A local Large Language Model (LLM), executed via Ollama and controlled through Lang Chain, performs generative and analytical tasks. Retrieval-Augmented Generation (RAG) mechanisms fetch relevant content segments from the vector database and supply them to the LLM for the production of summaries with keywords, question–answer pairs, flashcards, lecture plans, assignment items, and analytical insights. Additional modules process the same content to detect conceptual gaps and perform document comparison. All system functionalities are accessed through a Streamlit interface that enables document upload, tool selection, and output visualization.