EDUVISION: AN AI-POWERED PLATFORM FOR VISUALIZING, TRANSLATING, AND ENRICHING TEXTBOOK CONTENT

Project Reference No.: 49S_BE_2561

College : *Khaja Bandanawaz University, Kalaburagi*
Branch : *Department of Computer Science & Engineering*
Guide(s) : *Dr. Raafiya Gulmeher*
Student(s) : *Ms. Harmeen Safoora*
Ms. Saniya Saba
Mr. Syed Sameer
Mr. Mohammed Hamza

Keywords:

Artificial intelligence in education, Vision-language models, Educational document analysis, Automated text summarization, Multilingual content translation, Cognitive load reduction, Multimedia learning theory, Adaptive learning technologies.

Introduction:

PDF documents are widely used in education for distributing textbooks, lecture notes, and references. Despite their structural advantages, PDFs remain static and require significant manual effort for key information extraction, diagram interpretation, and summarization. The rise of Large Language Models (LLMs) has opened new possibilities for automating knowledge extraction and enhancing digital learning. However, general-purpose AI tools lack a tailored approach for structured academic PDF processing and are prone to context limitations, hallucination, and inconsistent output. EduVision addresses these challenges by providing a modular AI-powered pipeline that extracts text and images from uploaded PDFs, generates structured bullet-point summaries using Azure OpenAI GPT-4o, supports multilingual translation, identifies anatomical organs in medical images via vision AI, and maps them to pre-validated static anatomical diagrams. Unlike conversational AI assistants, EduVision uses separate FastAPI endpoints for each function, ensuring controlled, reliable, and educationally meaningful outputs grounded in Cognitive Load Theory and Multimedia Learning principles.

Objectives:

1. **To perform summarization** by extracting and condensing key information from uploaded PDFs or textbooks.
2. **To generate images** that are contextually relevant to the extracted content, or to identify suitable visuals already present within the PDF.
3. **To identify prerequisite knowledge** by detecting key terms and concepts that require foundational understanding before engaging with advanced material.
4. **To translate** content by converting generated summaries into the learner's preferred language

Methodology:

The proposed system leverages **Azure AI Foundry** to build and deploy AI-powered agents for summarization, translation, and content enhancement. It integrates **GPT-4o**, **Azure Cognitive Services**, and a **Flask-based web application** to process educational PDFs efficiently.

1. **Text Summarization:**
   The system uses **FastAPI-based backend system** to handle file uploads and route user requests. **PyMuPDF** extracts text, images, and links from PDFs, which are then divided into smaller text chunks to maintain relevance during summarization. **GPT-4o** generates concise, bullet-point summaries and ensures contextual coherence. Summaries are stored in JSON format and displayed via a simple user interface.
2. **Image Generation:**
   Azure Computer Vision extracts and analyzes images from the uploaded content using OCR. To support visual learning, The system maps detected anatomical structures to pre-validated static images instead of generating images, ensuring accuracy and avoiding hallucination. **Vision APIs** further assist in captioning and translating images for accessibility.
3. **Prerequisite Links:**
   Using **Azure Cognitive Search** and **Language Services**, the system identifies key concepts from study materials and links them to relevant online resources. This helps learners understand prerequisite topics before proceeding to the main content.
4. **Language Translation:**
   Azure Translator, built on transformer-based models, converts summaries and prerequisite content into the learner's preferred language. It also translates embedded text within images via OCR, ensuring inclusivity for learners from diverse linguistic backgrounds.

Overall, the methodology integrates Azure's AI ecosystem with FastAPI-based backend system processing to provide a seamless, intelligent, and accessible e-learning platform. It is not a theoretical or simulation-only study, but a functional prototype capable of processing real-world educational PDF documents.

Result and Conclusion:

EduVision successfully transforms static PDF documents into structured, accessible, and interactive learning material. Testing with academic and medical PDFs produced consistent bullet-point summaries, anatomical organ identification, detailed explanations, and multilingual translations. Key outcomes include: significantly reduced reading time via AI-generated summaries; improved concept clarity through vision-based organ identification and static anatomical diagrams; reduced cognitive load through chunked, structured outputs aligned with Cognitive Load Theory and Dual Coding Theory; and enhanced accessibility for non-English speakers through

GPT-4o translation. The static image mapping approach eliminated hallucinated diagrams, ensuring medical accuracy. EduVision's modular FastAPI architecture with separate endpoints ensures clean separation of concerns, easy maintenance, and scalable extension. The system demonstrates that AI-powered tools can meaningfully enhance digital learning by bridging the gap between static content and intelligent, adaptive educational experiences.

EduVision has strong industry relevance in the fields of EdTech, digital publishing, and AI-powered learning platforms. It can be integrated into Learning Management Systems (LMS), e-learning platforms, and academic content providers to automate content summarization, enhance accessibility, and improve student engagement. The system addresses real-world challenges such as information overload, lack of multilingual support, and difficulty in understanding complex visual content, making it highly applicable in modern education ecosystems.

Key learnings from this project include practical experience in integrating Large Language Models and vision models into a real-world application, designing modular AI pipelines, and implementing structured prompt engineering for reliable outputs. The project also provided insights into handling AI limitations such as hallucination and context constraints, while applying educational theories like Cognitive Load Theory to improve system design.

References:

- S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran, and Z. Du, "Artificial Intelligence in Education: A Systematic Literature Review," 2024.
- K. Ouwehand, F. Lespiau, A. Tricot, and F. Paas, "Cognitive Load Theory: Emerging Trends and Innovations," 2025.
- Z. Ji, N. Lee, R. Frieske et al., "Survey of Hallucination in Natural Language Generation," 2024 (Updated).
- P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, "A Systematic Survey of Prompt Engineering in Large Language Models," arXiv, 2025.
- V. Heilala, R. Araya, and R. Hämäläinen, "Beyond Text-to-Text: Multimodal and Generative AI for Education," 2024.

Future Scope:

The future scope of this project includes:

1. Automated quiz and assessment generation from summarized content, enabling learners to test their understanding interactively and reinforcing knowledge retention.
2. Integration of voice-based interaction through Text-to-Speech and speech recognition, enhancing accessibility for visually impaired learners and those with limited reading ability.
3. Integration with Learning Management Systems (LMS) for institutional adoption, and personalization of summaries and visuals based on individual learner behavior and performance analytics.
4. Personalized learning according to learner behaviour or performance analytics.