

RAG-Based AI Placement Assistant

College placement cells manage a fragmented information landscape: eligibility policies, company visit archives, package records and frequently asked questions scattered across PDFs, spreadsheets and notice boards. Students repeatedly query coordinators for information that exists but is hard to retrieve, while generic large language models cannot help — they lack institutional context and hallucinate specifics such as eligibility criteria or salary figures. This project presents PlaceMate, a retrieval-augmented generation (RAG) chatbot that grounds a large language model in the placement cell's own document corpus.

The system implements a modular ingestion pipeline: uploaded documents (PDF, DOCX, webpages) are parsed, cleaned and split into overlapping chunks using a recursive text splitter, then converted to dense vector embeddings and indexed in a vector database (FAISS/Chroma). At query time, a LangChain retrieval chain performs semantic similarity search to fetch the top-k relevant chunks, which are injected as grounded context into the LLM prompt alongside conversation history for follow-up resolution. Generation uses OpenAI-compatible or Gemini-compatible LLM APIs, and every answer carries citations to its source documents and page references.

Hallucination control is architectural rather than prompt-deep: the generation chain is constrained to retrieved context, and queries with no relevant matches receive an explicit "not found in documents" response instead of invented answers. The frontend is a Streamlit application with a chat interface, source citation viewer, chat history and an admin upload panel. Evaluation covers retrieval precision over a labelled placement FAQ set, citation accuracy and response latency. The result is a demonstrable, viva-ready LLM system that converts unstructured institutional knowledge into a trustworthy, source-cited conversational assistant.