

# RETRIEVAL-AUGMENTED GENERATION SOLUTION POC FOR AN AERONAUTICAL UNIVERSITY

THE APPLICATION ENABLES USERS TO EFFICIENTLY QUERY AND  
ANALYZE DATA FROM PDF REPORTS

## CUSTOMER PROFILE

### HQ

Daytona Beach, FL

### INDUSTRY

Higher Education

### EMPLOYEES

2,500

### ITC SERVICES

- Artificial Intelligence and Machine Learning

### APPLICATIONS & TECHNOLOGIES

- Llama 3.1
- LangChain
- Streamlit

## INTRODUCTION

The client is a prestigious institution specializing in aviation, aerospace, and related fields. They are recognized globally for offering high-quality education and training, with a strong focus on science, engineering, and technology in these industries. The university offers courses in areas such as aeronautical science, aerospace engineering, unmanned aircraft systems, aviation business, and air traffic management.

## CHALLENGES

The client had a collection of documents and incident reports related to flights and aviation in PDF format that they wanted to query and extract insights from. The challenge is that, despite having all this information, they were unable to effectively use or analyze the data. They wanted to conduct in-depth analysis and leverage this information to gain valuable insights and make better use of their data.

## SOLUTIONS

IT Convergence proposed building a Retrieval-Augmented Generation (RAG) application using open-source technology with the client's own data. We developed a proof of concept for a RAG application by utilizing an open-source LLM from Meta to answer user questions based on the data in the PDF reports. The PDF files were ingested into an open-source vector database to allow the LLM to process them. However, the metadata of the reports was not well-organized, making it difficult for the LLM to interpret the information accurately, so we recommended improving the metadata for the final project. To facilitate interaction with the data, we created a chat interface using open-source software, allowing users to ask questions about their data. The LLM then responds to these questions based on the data stored in the vector database.

## RESULTS

- The application quickly retrieves information from large data sets, allowing users to ask questions and receive instant answers from PDFs without the need for manual searches. This saves time and increases efficiency.
- It provides responses that cannot be obtained manually by overcoming the limitations of the unstructured nature of the original data sources.
- By combining document retrieval with generation, the application delivers more accurate and detailed responses than simple keyword searches or manual PDF reviews. It can uncover insights and patterns that would be difficult to detect through manual methods.
- Automating the retrieval and analysis of PDF content eliminates time-consuming manual searches, enabling better data-driven decisions based on historical reports.

## ITC ADVANTAGE

- Deep knowledge of open-source technologies to build customized RAG applications, ensuring cost-effective and scalable solutions
- Expertise in integrating diverse data sources and advanced analysis, enabling actionable insights from complex and large-scale datasets