Smart PDF Insight Generator
Business Technical Report Demo

Konan Othniel

June 2025

Abstract

This document is used to demonstrate the automatic extraction and summarization capabilities of the **Smart PDF Insight Generator**. It includes business context, technical details, and formatted content to test AI summarization.

Contents

1 Introduction

Companies handle hundreds of PDF documents daily—contracts, invoices, reports, scientific papers. Manually extracting insights from these files is slow, error-prone, and inefficient.

2 Objectives

- Automatically extract text from any PDF.

- Generate concise, actionable summaries using Large Language Models (LLMs).

- Demonstrate robust handling of English business documents.

- Support key features: lists, tables, equations, hyperlinks.

3 Methodology

3.1 Architecture Overview

- **AWS Lambda**: serverless backend for processing.

- **Groq LLM API**: advanced language model for summarization.

- **Postman/Python Client**: submits PDFs as base64 for testing.

3.2 Sample Equation

The total number of possible summary combinations, C, for n sections is given by:

$$C = 2^n - 1$$

3.3 Sample Table

Section	Content Type	AI Handling
Introduction	Text	Yes
Objectives	Bullet List	Yes
Methodology	Table, Equation	Yes
Results	Paragraph, Link	Yes
Conclusion	Text	Yes

Table 1: Example content types and AI processing

3.4 Sample Hyperlink

For more information, visit the GitHub repository: `https://github.com/Matheux14/smart-pdf-insight-gen`

4 Results

After submitting this PDF, the AI system should be able to:

- Extract all text, lists, and table data.

- Identify the main goals and findings.

- Produce an English executive summary for business users.

5 Conclusion

The Smart PDF Insight Generator enables companies to unlock knowledge hidden in their documents. It is robust, language-aware, and ready for real-world business workflows.