

# Sample Document for Testing

## 1. Introduction

This is a sample PDF document created for testing the document processing pipeline in LightRAG. It contains text, tables, and structure that can be used to test various features of the system.

## Table of Contents

1. Introduction..... 1  
2. Features..... 1  
3. Sample Table..... 2  
4. Conclusion..... 2

## 2. Features

LightRAG provides several features for document processing and knowledge extraction:

- Advanced PDF parsing with structure preservation
- Content filtering to remove headers, footers, and non-informative content
- Table extraction and conversion to Markdown
- Diagram and formula detection with placeholder replacement
- Schema-driven Knowledge Graph construction

## 3. Sample Table

The following table shows sample data that can be extracted by the table extraction component:

Model	Type	Parameters	Performance
GPT-4	Transformer	1.8T	High
Claude	Transformer	Unknown	High
Llama 2	Transformer	70B	Medium
Mistral	Transformer	7B	Medium

## 4. Conclusion

This sample document demonstrates various features that can be processed by the LightRAG document processing pipeline. It includes structured text, a table of contents, a data table, and multiple sections with headings.

The document processing pipeline should be able to extract the main content, filter out the table of contents, and extract the data table for further processing.