

Test Document for PDF Processing

This is a test document created for validating PDF processing functionality.

It contains multiple pages with different types of content to test:

- Text extraction capabilities
- Multi-page processing
- Various text formatting
- Paragraph and line breaks

This document should be processed by the PDF chunking system to extract meaningful text blocks and validate the extraction pipeline.

Chapter 1: Sample Content

This page contains sample text that mimics real document content.

It includes paragraphs, dialogue, and various text patterns.

"Hello, this is a sample dialogue," said the first speaker.

"Yes, this helps test dialogue detection," replied the second.

Regular paragraph text continues here. This text should be processed as a normal paragraph block, separate from the dialogue above.

Another paragraph with some technical terms like PyMuPDF4LLM and text processing algorithms to test specialized handling.

Chapter 2: Structured Content

This page tests structured content processing:

1. First numbered item
2. Second numbered item
3. Third numbered item

Bullet points:

- First bullet point
- Second bullet point
- Third bullet point

This content tests the system's ability to handle different text structures and formatting patterns.