Comprehensive Formatting Test Document

This document contains various markdown formatting elements to test PDF extraction capabilities.

Text Formatting

This paragraph includes **bold text**, *italic text*, ***bold and italic***, ~~strikethrough~~, and `inline code`.

Lists

Unordered List

- First item
- Second item with **bold**
  - Nested item 1
  - Nested item 2
    - Deep nested item
- Third item

Ordered List

1. First step
2. Second step
   i. Sub-step A
   ii. Sub-step B
3. Third step

Task List

- [x] Completed task
- [ ] Incomplete task
- [x] Another completed task

Code Blocks

Inline Code

Use the `extract_text_from_pdf()` function to process PDFs.

Code Block with Syntax Highlighting

```
pub async fn extract_text_from_pdf(path: &Path) -> Result<PdfContent, CliError> {
    match extract_with_pdfium(path).await {
```

```
        Ok(content) => Ok(content),
        Err(_) => extract_with_pdf_extract(path).await,
    }
}
```

```
{
  "status": "success",
  "page_count": 42,
  "extractor": "pdfium-render"
}
```

Tables

Simple Table

Column 1	Column 2	Column 3
Row 1-A	Row 1-B	Row 1-C
Row 2-A	Row 2-B	Row 2-C
Row 3-A	Row 3-B	Row 3-C

Table with Alignment

Left Aligned	Center Aligned	Right Aligned
Left	Center	Right
A	B	C
100	200	300

Complex Table

Feature	Description	Status	Priority
PDF Extraction	Dual-strategy extraction	✅ Complete	High
Input Guardrails	PII & prompt injection detection	✅ Complete	High
Output Guardrails	Safety & quality checks	✅ Complete	Medium
REST API	Actix-based HTTP endpoints	✅ Complete	Medium

Links and References

External Links

- Anthropic
- Rust Programming Language
- GitHub Repository

Internal Links

- Jump to Text Formatting
- Jump to Tables
- Jump to Blockquotes

Reference-style Links

Visit Anthropic or check out Rust.

Blockquotes

> This is a simple blockquote.

> This is a multi-line blockquote. It spans multiple lines. And contains **formatted text**.

> ### Blockquote with heading
>
> Blockquotes can contain other elements:
>
> - Lists
> - **Bold text**
> - Code

> Nested blockquote:
>
> > This is nested
> >
> > > And this is deeply nested

Horizontal Rules

Images (Reference)

![Alt text for image]

![Logo]

Special Characters and Escaping

Special characters: & < > " ' @ # $ % ^ * ( ) [ ] { } | \ / ? ! ~ `

Escaped characters: * _ # [ ] ( ) ` \

Footnotes

Here's a sentence with a footnote.[^1]

Another reference to a footnote.[^2]

[^1]: This is the first footnote. [^2]: This is the second footnote with **formatting**.

Definition Lists

Term 1

Definition 1a

Definition 1b

Term 2

Definition 2

Abbreviations

The HTML specification is maintained by the W3C.

*[HTML]: Hyper Text Markup Language *[W3C]: World Wide Web Consortium

Mathematical Expressions

Inline math: $E = mc^2$

Block math:

```
∫₀^∞ e^(-x²) dx = √π/2
```

Emojis and Unicode

Emojis: 🚀 ✅ ❌ ⚠️ 📄 💡 🔒 💥

Unicode: α β γ δ ε ζ η θ → ← ↑ ↓ ∞ ∑ ∏ √∫

Long Paragraph Test

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. This paragraph tests the extractor's ability to handle long continuous text without line breaks and should maintain readability in the extracted output.

Mixed Content Section

This section combines **multiple formatting types**:

1. **Bold numbered list item** with `inline code`
2. *Italic item* with [a link](#)
3. Normal item with ~~strikethrough~~

Followed by a table:

Metric	Value	Unit
Latency	150	ms
Throughput	1000	req/s
Error Rate	0.1	%

And a code block:

```
cargo test --all
cargo run --bin llm-eval-api-server -- --port 8080
```

> And a blockquote with a list:
>
> - Item A

- Item B
- Item C

Nested Structures

Level 3 Heading

Level 4 Heading

Level 5 Heading

Level 6 Heading

Regular paragraph after deep nesting.

HTML Elements (if supported)

Some **HTML** content

- HTML list item 1
- HTML list item 2

▶ Click to expand

Final Section

This is the final section of the comprehensive test document. It should appear on page 3 or 4 depending on export settings.

Summary Statistics

- **Total Sections**: 20+
- **Formatting Types**: 15+
- **Test Coverage**: Comprehensive
- **Expected Page Count**: 3-5 pages

End of Document

Last updated: 2026-01-12