

Intentionally Problematic PDF (Page 1)

This PDF is intentionally constructed with layout/content issues for testing parsers/viewers.

1) Text covered by other elements

This sentence should be partially covered by a red



2) Hidden text (white-on-white and very small)

Below is 'missing' text: it exists in the PDF but is invisible or tiny.

THIS IS A VERY SMALL AND HARD TO READ COMMENTARY LINE

3) Overlapping / duplicated text

OVERLAP TEST: These lines overlap heavily.

4) Clipped text (only part shows)

This long line is clipped by a small clipping box.

5) Text placed off-page (exists but not visible)

Some text is drawn outside the page bounds (negative coords / beyond width).

Intentionally Problematic PDF (Page 2)

This PDF is intentionally constructed with layout/content issues for testing parsers/viewers.

6) Large image blocks underlying text



7) Low-contrast text (hard to read)

This text is intentionally low-contrast against white background.

8) Rotated text overlapped by shapes

ROTATED: This should be partially covered by a dark box.

BOX ON TOP

Intentionally Problematic PDF (Page 3)

This PDF is intentionally constructed with layout/content issues for testing parsers/viewers.

9) Potential missing glyphs (unsupported characters)

Text with emojis/symbols: 🐼 🐼 🐼 🐼 🐼 (some may render as blank/tofu depending on font/viewer)

Combining marks: aaa éée Žálgó

Background band. Next line uses alpha if available; otherwise it's light grey on grey.
FAINT TEXT: alpha=0.05 (very hard to see)

11) Form field (may not extract like normal text)

Below is a text field annotation: some parsers ignore it or treat it differently.

12) Mixed reading order (columns with crossing z-order)

Left column line 1

Left column line 2

Left column line 3

Right column line A

Right column line B

Right column line C