OCR ONE-PAGE FULL TEXT TEST DOCUMENT

This document is designed to test OCR text extraction, noise reduction, emoji handling, and layout accuracy in a single full page. It contains normal sentences, emojis ■ ■ ■ ■ ■, mixed symbols, numbers, and intentionally noisy text.

Normal Text Block: OCR systems should correctly extract this paragraph without losing words or punctuation. The quick brown fox jumps over the lazy dog. This sentence is repeated to test consistency and accuracy. The quick brown fox jumps over the lazy dog again.

Emoji Test: ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■

Noise & Symbols Test: @@@ ### $$$ %%% ^^^ &&& *** --- ___ !!! ??? OCR should either clean or preserve these based on configuration.

Broken & Noisy Text: Th1s t3xt c0nta1ns n0is3, br0k3n w0rds, rand0m numb3rs 12345, and symb0ls!!! Use this section to validate noise reduction logic.

Layout Order Check: Left-side logical flow → middle continuation → right-side ending. If OCR reads out of order, the extracted text sequence will not match this sentence. This helps validate layout handling.

Tabular-like Inline Data: ID:1 Status:Processed■ File:Invoice_001.pdf | ID:2 Status:Failed■ File:Resume(3).pdf | ID:3 Status:Processed■ File:Form_A12.pdf

Final Check Paragraph: This is the end of the one-page OCR test document. If all text above appears correctly in your extracted gtext output, then emoji handling, noise processing, and layout detection are working as expected.