REALTEST.ORG SCANNED PDF OCR BENCHMARK V1.0

DOCUMENT TYPE: 150 DPI RASTER SCAN ONLY | EMBEDDED PDF FONTS: 0 | PAGES: 1 OF 2

SECTION 1: OPTICAL CHARACTER RECOGNITION & TEXT LAYER AUDIT

Scanned PDF documents contain zero native font streams and require true optical character recognition to recover searchable text, invoice identifiers, and structured financial numbers. Tools that only extract existing PDF text layers without running an OCR engine produce blank output.

SECTION 2: INVOICE & FINANCIAL TABLE OCR ACCURACY

Invoice ID	Amount USD	Confidence	Verification Code
INV-2026-0108	$128,450.00	99.42%	OCR-ALPHA-01
INV-2026-0219	$256,900.50	98.85%	OCR-BETA-02
INV-2026-0387	$512,750.25	99.91%	OCR-GAMMA-03

CONFUSABLE GLYPH AUDIT BLOCK (0 vs O, 1 vs l vs I, 5 vs S, 8 vs B):

BATCH-ID: RT-2026-8051 | SERIAL: 0192-8374-5500 | STATUS: VERIFIED-CLEAN

SECTION 3: PAGE 2 MULTI-PAGE SCAN CONTINUATION

Page two verifies that free OCR tiers process multi-page scanned documents without truncating after page one or locking searchable PDF downloads behind undisclosed post-OCR paywalls.

END-TO-END MULTI-PAGE OCR VERIFICATION TOKEN:

REALTEST-PDFOCR-PAGE2-VERIFIED-2026

SUMMARY OF EXPECTED OCR DELIVERABLE PROPERTIES:

1. Searchable PDF (.pdf) overlays an invisible selectable text layer aligned with the scanned page image.
2. All 4 section headings and all 12 invoice table cells must be recoverable via standard text selection.
3. Multi-page completeness requires preserving both Page 1 financial rows and the Page 2 token above.