

Local Detection Hardening Report

A failed-versus-improved comparison of deterministic OCR candidate generation across seven controlled pages and four scan conditions, scored against labeled sensitive regions and hard-negative page text.

VERDICT	PASS - ALL DECLARED GATES CLOSED
BOUNDARY	This is a controlled synthetic English corpus with photometric scan perturbations. It does not establish uncontrolled-document, handwriting, face, multilingual, or universal PII recall.
CORPUS	7 pages x 4 variants = 28 runs; 104 labeled text-region instances
RUN ID	HEAD 31a4c7c66a8b / STATUS ec1fbb6b5b8e / DIRTY

84.62% -> 100.00% REGION RECALL	81.48% -> 100.00% CANDIDATE PRECISION	99.98% CURRENT TARGET GLYPH COVERAGE	0.000% BENIGN LOSS OUTSIDE HALO
------------------------------------	--	---	------------------------------------

Performance by Scan Condition

VARIANT	REGIONS	CANDIDATES	RECALL	PRECISION	GLYPH COVERAGE
original	26	26	100.00%	100.00%	99.98%
low-contrast	26	26	100.00%	100.00%	99.99%
blur-1	26	26	100.00%	100.00%	99.98%
scan-noise	26	26	100.00%	100.00%	99.98%

The same labeled regions were evaluated on original raster pages, reduced-contrast pages, a deterministic one-pixel blur, and deterministic scan noise. These are correlated variants used for robustness testing, not independent documents.

What the Failed Run Taught Us

REGION TYPE	ITERATION 1 MISSES	CURRENT MISSES	MECHANISM
name	16	0	Label punctuation, Subject labels, and Holder-column propagation.

Iteration 1 produced 20 false-positive candidates and missed 16 labeled regions. The current run produced 0 false positives and missed 0 regions. Title-like text and field labels are now rejected unless punctuation or value structure supplies evidence.

Dashed Truth Versus Solid Candidates

text-pii / PAGE 1 / BLUR-1 / 6/6 RECALL

mixed / PAGE 3 / BLUR-1 / 6/6 RECALL

Customer Verification Record

CONTROLLED SYNTHETIC FIXTURE

All values below are synthetic and intentionally planted for recovery testing.

Full name: Avery Jordan

SSN: 017-92-9088

Email: avery.jordan@example.test

Phone: +1 415 555 0198

Home address: 1847 Cedar Lane, Oakland, CA 94612

Account number: ACC1-8841-9927-3308

Review status: pending redaction approval

Settlement Table

PAGE 3 - DIGITAL TABLE

Account	Holder	Balance
ACC-447138200	Leah Stone	\$4,120.00
ACC-447138200	Omar Reed	\$8,811.40
ACC-447138202	Lara Kim	\$2,006.18

DASHED = LABELED SENSITIVE REGION. SOLID = LOCAL OCR CANDIDATE

The overlay is generated from the benchmark JSON. Dashed boxes are labeled sensitive regions; solid boxes are local OCR candidates. The visual remains secondary to the machine-scored one-to-one region matching and target-glyph coverage checks.

Useful Now, Bounded Honestly

GATE	THRESHOLD	MEASURED	STATUS
recall	minimum 90.000%	100.000%	PASS
precision	minimum 90.000%	100.000%	PASS
targetGlyphCoverage	minimum 99.000%	99.983%	PASS
minimumVariantRecall	minimum 90.000%	100.000%	PASS
minimumVariantPrecision	minimum 90.000%	100.000%	PASS
minimumVariantGlyphCoverage	minimum 99.000%	99.980%	PASS
benignGlyphLoss	maximum 0.100%	0.000%	PASS

Product value: when GPT or SAM credentials are absent, the user still receives deterministic candidate coordinates for common labeled English PII instead of an empty manual-mode response. The release remains reviewer-gated, and the 12px destructive overwrite path is evaluated separately from detector recall.

Not claimed: universal PII detection, face detection, handwriting recognition, multilingual recall, legal compliance, or safety without human review. The next evidence tier is a labeled uncontrolled corpus with real scans and independent document sources.
