

Sample Hebrew OCR Defect Analysis

Intended use: synthetic invoice OCR and field-extraction review.

Source status: Synthetic public demonstration. No customer data is included.
Limitations: This sample demonstrates documentation structure only. Project-specific rules, thresholds, and outputs must be agreed separately.

Scope

A synthetic set of 40 Hebrew invoice images was reviewed for OCR text, reading order, table structure, and critical field extraction.

Key results

Metric	Result	Target
Word accuracy	93.8%	≥ 97%
Invoice number extraction	95.0%	≥ 99%
Total amount extraction	92.5%	≥ 99%
Table cell accuracy	88.4%	≥ 95%
Reading-order accuracy	90.0%	≥ 98%

Defect examples

Category	Example	Business impact
RTL reading order	Column labels were read after the values.	Incorrect field pairing
Digit confusion	8 read as 3 in the total amount.	Critical financial error
Table merge	Two adjacent line items were combined.	Loss of item-level detail
Mixed language	English SKU separated from Hebrew description.	Broken item matching

Recommended actions

- Add more low-contrast and mobile-captured invoices to the test set.
- Evaluate critical fields separately from general OCR accuracy.
- Introduce a post-extraction validation rule for totals and currency.
- Track reading-order failures as a distinct defect category.