

SOLUTION DOSSIER

AI Document Processing

OCR and AI extraction for multilingual invoices, receipts, contracts, scans and emails, producing structured output with confidence scoring and a searchable archive.

apporix.lu/solutions/ai-document-processing-luxembourg

What separates this from basic OCR

- **Luxembourgish extraction from research** — For Luxembourgish documents, Apporix follows and uses recent scientific advances from the University of Luxembourg in language technology. Extraction is designed for French, German, English and Luxembourgish content, including documents that mix them.
- **Confidence scoring before anything is committed** — Low-confidence fields, unclear scans and unusual layouts are flagged for human review before records are updated or a document moves to the next step. Accuracy depends on scan quality, language, layout variety and field complexity, so uncertain and high-risk fields keep human validation rather than being assumed correct.
- **Structured output into real systems** — File content becomes clean fields for Excel, CSV, JSON, CRM, ERP, Odoo, BOB50, databases or dashboards — not a text blob requiring re-entry.
- **Searchable archive with embeddings** — Processed documents are stored with metadata, embeddings and traceable context, so approved files can be retrieved quickly later.
- **Sensitivity-scaled handling** — Access rights, logging, data minimisation and human review are designed around how sensitive the files are, with GDPR, EU data residency and zero-retention options where required.

Document types handled

Invoices, receipts, contracts, customer requests, onboarding forms, documentation, notes, reports, certificates, delivery notes, and emails with attachments. Handwriting and poor-quality images can be processed but usually need additional validation.

How a project runs

01	Collect samples — Review real invoices, receipts, contracts, scans, notes and edge cases to understand the actual formats.
02	Define fields — Set the fields, OCR needs, confidence rules, validation checks and structured outputs.
03	Prototype — Test extraction, classification and document search on realistic batches before production.
04	Integrate — Connect outputs to email, storage, CRM, ERP, Odoo, BOB50, databases or dashboards.
05	Improve — Review errors, refine extraction rules and expand document types once validation is reliable.