



Intelligent document processing

Extracting, classifying, validating, and routing information from business documents.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 06 / EMBER & LIME

EXECUTIVE WORKING EDITION - 08 OCT 2026

06

AT A GLANCE

Extracting, classifying, validating, and routing information from business documents.



Intelligent document processing (IDP) turns invoices, contracts, forms, claims, certificates and correspondence into structured data that systems and people can act on. It combines reading the page, classifying the document, extracting fields, validating them against rules and routing exceptions to a person. Modern models read messy layouts far better than older template tools, but they still make errors that look confident. This playbook explains where IDP fits, how the pipeline works, who owns what, how to set confidence thresholds and review rules, how to measure accuracy by field, and how to move from a single document type to a production process.

WHO IT IS FOR

Operations, finance, claims, procurement, compliance and technology leaders who own document-heavy processes, and the process owners and reviewers who will run them every day.



TIME TO USE

70 minutes to read; two workshops to apply (document inventory and accuracy test)

What you will have at the end

- A ranked inventory of document types by volume, effort and risk
- A field-level accuracy target and a confidence-threshold policy for one document type
- A validation and exception design that keeps a named person accountable
- A test set and measurement plan that reports accuracy by field, not just overall
- A 90-day plan from one document type to a go, change or stop decision



Contents

01	What intelligent document processing means for the business	03
02	Where IDP fits and where it does not	04
03	How the pipeline works, in plain language	06
04	Operating model, roles and exception ownership	08
05	Risks, controls and human review	09
06	Measuring accuracy, flow and value	11
07	Implementation guide: one document type, end to end	13
08	Common pitfalls and an illustrative scenario	15
09	Questions to ask vendors, and build or buy	16
10	A first quarter that produces evidence	18
11	Terms used in this publication	19



EXECUTIVE BRIEF

What intelligent document processing means for the business

01



IDP is not about reading documents faster. It is about getting correct data into the right system with a clear route for everything the machine is unsure about.

Most organizations run on documents that arrive as PDFs, scans, photos, emails and spreadsheets. Someone opens each one, finds the supplier name, the amount, the date, the policy number or the clause, and keys it into another system. That work is slow, repetitive and error-prone, and the delays ripple into payments, claims, onboarding and reporting. IDP automates the reading and the first pass of checking, so people spend their time on the cases that need judgment.



Older tools needed a template for each layout and broke when a supplier changed its invoice design. Current approaches use language and vision models that interpret a page more like a person does, including layouts they have never seen. That is a real improvement, but it shifts the risk. A template tool fails loudly when it cannot find a field. A model may return a plausible value that is wrong, such as a total taken from the wrong line. The design of validation and review therefore matters more than the choice of extraction engine.



- The unit of value is a correct field in the downstream system, not a document that was processed
- Accuracy must be reported field by field, because a 98 percent document score can hide a field that is wrong one time in five
- The goal is straight-through processing for clean, low-risk documents and fast review for everything else
- Exceptions are not a failure of the system; they are the part of the design that protects the business
- Document variety, scan quality and downstream rules decide the outcome more than the model does



PLAIN TEST

If a trained clerk reads the same fields from the same kinds of documents every day, and a second person could check them against a rule or a source system, IDP probably fits. If the work depends on reading intent between the lines, treat it as analysis, not extraction.

WORKING QUESTION

Which single document type costs you the most clerk hours each month, and what happens downstream when a field on it is wrong?

WHERE IT FITS

Where IDP fits and where it does not

02

IDP pays off when documents are frequent, fields are well defined and errors can be caught by rules. It disappoints when the real problem is a broken process or missing upstream data.

IDP sits among several options. Electronic data interchange and supplier portals avoid documents altogether by receiving structured data at the source, and they are often better where trading partners will cooperate. Classic optical character recognition (OCR) with templates still works for fixed forms that never change. Generative models handle variable layouts and long free text, and can summarize or answer questions about a document. Many real solutions combine these: OCR or layout parsing to read the page, a model to interpret it, and rules to check it.

APPROACH	BEST WHEN	WATCH FOR
Structured data at source (portal, e-invoicing, API)	Partners can send data directly	Adoption effort; not every partner will comply

APPROACH	BEST WHEN	WATCH FOR
Template OCR	Fixed forms with stable layouts	Breaks when layouts change; high maintenance
Model-based extraction	Variable layouts, many senders, mixed quality	Confident wrong values; needs validation
Document question answering	People need to ask about long documents	Not a substitute for field-level extraction
Manual keying with checks	Very low volume or very high stakes	Cost and delay grow with volume

USE CASE	FIT	REASON
Supplier invoices matched to purchase orders	Strong	High volume, clear fields, match rules catch errors
Claims intake forms and attachments	Strong with review	Mixed formats; classification and routing add large value
Identity and certificate checks for onboarding	Moderate	Useful for extraction; authenticity checks need specialist controls
Contract clause extraction	Moderate	Good for locating terms; interpretation needs a professional
Handwritten forms of poor quality	Weak to moderate	Accuracy varies widely; test on real samples first
Documents that are rarely seen and all different	Weak	No pattern to learn; setup cost exceeds benefit

COMMON MISTAKE

Automating extraction while the process around it stays broken. If invoices are late because approvers are unavailable, faster data entry will not fix payment delays. Map the whole flow before choosing the target.

WORKING QUESTION

For your top document type, is the main delay reading the document, or waiting for someone to decide what to do with it?

HOW IT WORKS

How the pipeline works, in plain language

03

Five stages move a document from arrival to a record: capture, understand, extract, validate and route. Each stage has its own failure modes and its own control.

01 Capture

Receive documents from email, upload, scanner or system feed. Record the source, the time and the original file so every result can be traced back.

02 Pre-process and read

Clean the image, correct rotation, split multi-document files and convert the page to text and layout with OCR or a vision model.

03 Classify

Decide what the document is, such as invoice, credit note, delivery note or statement, because each type has different fields and rules.

04 Extract

Pull the defined fields, including line items, with a confidence signal for each and a pointer to where it was found on the page.

05 Validate and route

Check values against rules and reference data. Pass clean documents onward and send the rest to a review queue with the reason shown.

Several design choices carry business consequences. Classification errors cascade, because a credit note treated as an invoice sends the wrong sign into the ledger. Line items and tables are harder than header fields, and totals that look right can hide a missed row. Confidence scores are useful but not guaranteed to be calibrated, so a value reported as highly confident may still be wrong at a rate you must measure. Pointing each extracted value back to its location on the page lets reviewers confirm a field in a second instead of rereading the document.

STAGE	WHAT IT DOES	TYPICAL FAILURE
Capture	Collects files and metadata	Duplicates; attachments lost; email signatures read as documents
Reading	Turns pixels into text and layout	Skewed scans, stamps over text, faint print, handwriting
Classification	Names the document type	Credit note read as invoice; statement read as invoice
Extraction	Finds the defined fields	Wrong total chosen; tax and net swapped; line items merged
Validation	Checks against rules and sources	Rules too loose to catch errors or too strict to be usable
Routing	Sends to system or reviewer	Exceptions pile up with no owner; reasons not shown

KEY DISTINCTION

Reading errors and understanding errors need different fixes. If the text was misread, improve scan quality or the reader. If the text was read correctly but the wrong value was chosen, improve the instructions, the examples or the validation rules.

WORKING QUESTION

For one real document, can you say which stage would catch it if the total were taken from the wrong line?

OPERATING MODEL

Operating model, roles and exception ownership

04

IDP is a process with a technology inside it. It needs a process owner, trained reviewers and a routine for turning errors into improvements.

The most common operating failure is an exception queue nobody owns. Documents that the system cannot process pile up, deadlines pass and people go back to their old habits. Decide before launch who clears the queue, how quickly, and who is the backup. Treat the queue as a signal: a rising count for one supplier or one field means something upstream changed, and the fix may be a conversation with that supplier rather than a model change.

ROLE	OWNS	TYPICAL TASKS
Process owner	End-to-end outcome and service levels	Sets targets, approves rule changes, reports results
Document reviewer	Exception resolution	Confirms or corrects fields, records the reason
Data steward	Reference data	Maintains supplier lists, codes, tax rules, product masters
Rules and configuration	Validation logic and thresholds lead	Tunes rules, tests changes, keeps a change log
System owner	Pipeline and integrations	Monitors uptime, handles incidents, manages releases
Control owner (finance, risk or compliance)	Evidence and audit trail	Defines what must be retained and who may override

Review queue design

- Show the document image beside the extracted fields, with the suspect field highlighted
- Show why the item is in the queue: low confidence, failed rule, unknown sender or possible duplicate
- Capture the correction and the reason so patterns can be analyzed

- Set time targets per queue and an escalation path for items that age
- Separate quick fixes from items that need a call to a supplier or customer
- Rotate reviewers and sample their work so quality does not drift

Corrections are the most valuable data the process produces. Each fix shows where the system is weak. Review them monthly, group them by cause, and decide whether the answer is a better instruction, a new validation rule, a cleaner reference list or a change to what you ask suppliers to send.

WORKING QUESTION

If the exception queue doubled next week, who would notice, and who has the authority to act?

RISKS AND CONTROLS

Risks, controls and human review

05

The main risks are confident wrong values, silent drift, duplicate or fraudulent documents and exposure of sensitive data. Controls combine rules, sampling and clear human accountability.

Frameworks such as the NIST AI Risk Management Framework and ISO/IEC 42001 provide structure for governing AI systems. Data protection law applies when documents contain personal data, and sector rules, record-keeping duties and the EU AI Act may apply depending on the use. Documents often include bank details, identity numbers, health information and signatures. This playbook does not give legal advice. Confirm current requirements with qualified advisers and turn them into controls on access, retention, logging and review.

RISK	EXAMPLE	CONTROL
Confident wrong value	Tax amount captured as the total	Cross-field rules (net plus tax equals total); match to purchase order

RISK	EXAMPLE	CONTROL
Misclassification	Credit note processed as invoice	Classification confidence threshold; type-specific rules; sampling
Duplicate or fraudulent document	Same invoice submitted twice with changed number	Duplicate checks on supplier, amount and date; bank-detail change controls
Silent drift	Accuracy falls after a supplier changes layout	Weekly field-level monitoring; alert on exception rate shifts
Sensitive data exposure	Identity documents sent to a tool without agreed terms	Data classification; approved tools only; masking; retention limits
Prompt injection in documents	Text in a file instructs the model to approve payment	Treat document content as data; no actions triggered by document text
Over-reliance	Reviewers click approve without looking	Random deeper audits; measure reviewer override rates

Controls to confirm before launch

- Every extracted field is traceable to its location in the original document
- Validation rules exist for each critical field and are tested with bad examples
- Documents with authenticity or fraud concerns follow a separate specialist check
- Payment-related changes, such as new bank details, always require independent confirmation
- A named person approves overrides and the override is logged
- Retention, access and deletion rules match legal and policy requirements
- The process continues manually if the system is unavailable
- A sample of straight-through documents is reviewed every week

EXAMPLE (ILLUSTRATIVE)

An accounts payable team sets three tiers. Documents that match a purchase order and pass every rule are posted automatically. Those with one failed rule go to a clerk. Any change in bank details goes to a supervisor who confirms by calling a known number. The tiers follow risk, not convenience.

WORKING QUESTION

What is the single worst field to get wrong in your process, and which control would catch it before it reaches the ledger or the customer?



MEASURING VALUE

Measuring accuracy, flow and value

06



Measure accuracy by field and by document type, track how many documents flow through untouched, and count the cost of errors as well as the saving.

Begin with a baseline from the current process: documents per month, minutes of handling per document, rework rate, error rate by field and the time from arrival to posting. Then build a test set from real documents, not clean samples. Include poor scans, odd layouts, multiple languages if you operate in them, credit notes, duplicates, documents with missing fields and documents that should be rejected. Mark the correct values by hand. Aim for several hundred documents spread across senders, because a small set dominated by one supplier hides weaknesses.



METRIC	QUESTION IT ANSWERS	HOW TO MEASURE
Field accuracy	Is each field right?	Compare extracted values with the marked truth, per field
Document-level accuracy	Is the whole record right?	Share of documents with every critical field correct
Classification accuracy	Is the document type right?	Confusion table across types
Straight-through rate	How many documents need no touch?	Share posted without human action, with sampled checks
Exception rate and age	Is the queue healthy?	Count and ageing by reason and sender



METRIC	QUESTION IT ANSWERS	HOW TO MEASURE
Review override rate	Do reviewers correct the system?	Share of fields changed by reviewers
Cycle time	Is it faster end to end?	Arrival to posting, before and after
Cost of errors	What do mistakes cost?	Rework time, payment corrections, customer impact

Setting confidence thresholds

01 Score the test set

Record the confidence for each field and whether the field was correct.

02 Plot accuracy against confidence

Find the level above which accuracy meets your target for that field.

03 Set thresholds by field risk

Use a stricter threshold for amounts and bank details than for descriptive text.

04 Choose the review mode

Below the threshold, a person confirms. Above it, validation rules and sampling provide the check.

05 Revisit monthly

Thresholds drift as documents change, so re-test with fresh samples.

ILLUSTRATIVE FIGURES ONLY

Suppose a team handles 4,000 invoices a month at 6 minutes each, and the pilot lets 60 percent flow with a 2-minute check while the rest take 5 minutes. Handling falls from 24,000 to about 8,000 minutes. Deduct review of samples and fixes for errors before claiming the saving.

Report results to the process owner in plain terms: what percentage of fields were right, which ones were weakest, how many documents were posted without touch and what the errors cost. A single headline accuracy number invites false comfort. A field-level table invites the right conversations.

WORKING QUESTION

What accuracy would each critical field need before you would let documents post without a human looking?



IMPLEMENTATION GUIDE

Implementation guide: one document type, end to end

07



Start with one document type and one downstream outcome. Prove accuracy on real samples before adding types, senders or automation.

- 01 **Inventory documents**
List types, monthly volumes, senders, formats, handling time, error history and sensitivity. Rank by volume, effort and risk.
- 02 **Define the target record**
Specify each field, its format, whether it is required, the validation rule and the system of record that receives it.
- 03 **Collect and mark a sample**
Gather real documents across senders and quality levels and mark the correct values with two people checking a portion.
- 04 **Configure and test**
Set up reading, classification, extraction and validation. Run the sample, review failures by cause and adjust.
- 05 **Design review and routing**
Build the exception queue, thresholds, escalation and audit trail. Test it with the people who will use it.



06 Run in shadow mode

Process live documents alongside the manual process without acting on results. Compare field by field for several weeks.

07 Go live in stages

Start with review of every document, then move clean categories to sampled checks as measured accuracy supports it.

Readiness for one document type

- A process owner with authority over the workflow and the downstream system
- At least a few hundred real documents available for testing, with consent to use them
- Reference data, such as supplier lists and codes, is reasonably clean
- Integration route to the target system is understood, including error handling
- Legal and privacy review of the data in these documents is complete
- Reviewers are identified and have time reserved for the pilot
- A manual fallback is documented and tested

Allow for the work around the model. Cleaning supplier masters, agreeing field definitions with finance, adjusting purchase order practices and training reviewers commonly take more time than configuring extraction. Communicate with senders too: asking suppliers to send text-based PDFs instead of photographed pages, or to include a purchase order number, can improve accuracy more than any tuning.

Expand by adding one dimension at a time: a new sender group, a new language, a new document type. After each addition, re-run the test set and watch the weekly field-level report. Resist moving to full automation because a demonstration looked good. The evidence that justifies automation is weeks of stable, measured field accuracy on live documents.

WORKING QUESTION

Which document type would give you the clearest result in 60 days: high volume, stable layouts and a downstream rule that checks the data?

PITFALLS AND SCENARIO

Common pitfalls and an illustrative scenario

08



Most IDP programs stumble on scope, measurement and ownership, not on extraction technology.

- Testing on clean, vendor-chosen samples instead of your messiest real documents
- Reporting one overall accuracy figure that hides weak fields
- Ignoring line items until the pilot has already been declared a success
- Launching without an owner for the exception queue
- Setting thresholds once and never revisiting them as document mixes change
- Treating a model update or vendor change as invisible, then discovering accuracy moved
- Skipping duplicate and bank-detail controls because extraction accuracy looks high



ILLUSTRATIVE SCENARIO (HYPOTHETICAL)

A distribution business pilots invoice extraction and reports 97 percent accuracy. Finance finds that tax codes are wrong on one in six invoices from a few suppliers, which the overall figure had hidden. The team switches to a field-level report, adds a tax-code rule against the supplier master, routes the weak suppliers to review, and asks them to send text PDFs. Posting errors fall and the straight-through rate rises gradually.



PITFALL	EARLY WARNING SIGN	RESPONSE
Hidden field weakness	Reviewers keep fixing the same field	Field-level reporting; targeted rule or instruction
Queue backlog	Items older than the service target	Owner assignment; staffing review; fix top causes
Drift after layout change	Exception rate jumps for one sender	Alert by sender; retest; contact supplier
Reviewer fatigue	Override rate falls toward zero	Random audits with seeded errors



PITFALL	EARLY WARNING SIGN	RESPONSE
Scope creep	New document types added before the first is stable	Stage gates tied to measured results

WORKING QUESTION

If your overall accuracy looked excellent, which field would you check first to see whether it was hiding a problem?

BUYING QUESTIONS

Questions to ask vendors, and build or buy

09

Ask vendors to prove performance on your documents and to show how errors are found, explained and corrected.

VENDOR QUESTION	WHAT TO LOOK FOR
Can we test on our own documents, including poor scans, before we sign?	A structured trial with field-level results, not a curated demonstration
How do you report confidence and is it calibrated against accuracy?	Evidence from a test set, plus tools to set thresholds per field
How do you handle tables, lists, items and multi-page documents?	Specific performance results on line items, not only header fields
Can reviewers see where on the page each value came from?	Highlighted source locations and fast correction in the same screen
Where is data stored and processed, and is it used to train models?	Written terms, region options, retention and deletion controls

VENDOR QUESTION	WHAT TO LOOK FOR
What happens when you change a model or version?	Notice periods, version pinning and regression testing support
How do we export our configuration, corrections and audit trail?	Portability and no lock-in on labeled data
What integrations exist with our ERP, claims or case system?	Proven connectors and clear handling of errors and retries



Build or buy decision

- A packaged product handles your main document types and languages in testing
- You need custom handling for unusual documents or strict control over data location
- Your team can maintain rules, reference data and evaluation after launch
- Costs are modeled at pilot and production volumes, including review labor
- An exit plan exists if quality or terms change



Be cautious of claims that a product reaches a single accuracy figure across all document types. Accuracy depends on your senders, your scan quality and your definition of correct. Ask the vendor to commit to the test method, not to a number, and keep your own test set so you can compare options and re-test after any change.

WORKING QUESTION

If a vendor changed its model next month, how would you find out whether your accuracy changed, and who would do the checking?



30 / 60 / 90

A first quarter that produces evidence

10

Prove field-level accuracy on one real document type, then widen with evidence.

First 30 days: ground it

- Inventory document types and rank by volume, effort and risk
- Choose one document type and name a process owner
- Define the target record, fields and validation rules with finance or operations
- Collect and mark a sample of real documents, including poor ones
- Record the baseline for handling time, errors and cycle time

Days 31 to 60: test it

- Configure reading, classification, extraction and validation
- Run the test set and report accuracy by field and document type
- Set confidence thresholds by field risk
- Build the review queue, escalation path and audit trail
- Start shadow mode on live documents

Days 61 to 90: decide

- Compare shadow results with the manual process, field by field
- Go live with review of every document, then move clean categories to sampled checks
- Publish a weekly dashboard of field accuracy, exceptions and cycle time
- Decide to scale, change or stop; plan the next document type



GLOSSARY

Terms used in this publication

11

TERM	MEANING
Intelligent document processing	Reading documents, classifying them, extracting fields, validating the values and routing exceptions.
OCR	Optical character recognition, which converts images of text into machine-readable text.
Classification	Deciding what type a document is, such as invoice, credit note or claim form.
Field extraction	Pulling a defined value, such as total amount or policy number, from a document.
Confidence score	A signal of how sure the system is about a value; it must be tested against real accuracy.
Straight-through processing	A document passing from arrival to the target system with no human touch.
Exception queue	The list of documents sent to people because they failed a rule or fell below a threshold.
Validation rule	A check on an extracted value, such as net plus tax equals total.
Three-way match	Comparing an invoice, a purchase order and a goods receipt before payment.
Ground truth	The hand-verified correct values used to score extraction.
Shadow mode	Running the new process on live work without acting on its results, to compare with the current process.
Human in the loop	A design in which a person reviews or approves outputs at defined points.

A note on use

This publication is a public orientation resource, not legal, technical, financial or professional advice. Adapt it to your people, information, workflows and obligations, and check current tool terms, regulations and risks before acting.



TAKE THE NEXT STEP

Take the free Halden assessment at halden.company/assessments, or talk to us at halden.company/contact about turning this into a plan for your organization.

