



PRACTICAL GUIDE FOR SMBS, ACCOUNTING FIRMS AND ADMIN TEAMS

From PDF to Excel with AI

Learn to extract data from invoices, delivery notes, emails and PDFs with AI, and work out when doing it manually stops paying off.

CORE IDEA

Start with a manual, AI-assisted method. Automate when volume, errors or the need for traceability justify it.

This guide is for companies that receive documents on a recurring basis and want to turn them into useful data without replacing all their software.

This guide will help you	Expected outcome
Identify document processes worth automating	See where manual time is lost.
Test a simple flow with AI	Extract data without building software from day one.
Organise data in Excel	Work with a clean, reviewable table.
Calculate the cost of the manual process	Estimate whether automation can pay off.
Decide the next step	Keep it manual, improve Excel or build custom software.

A practical guide, not a sales brochure

Block	What you'll learn
1. Why this guide	The problem of documents that end up being copied by hand.
2. What you can automate	Invoices, delivery notes, emails, contracts, forms and complex PDFs.
3. Manual method with AI	How to test it with ChatGPT, Claude, Gemini or other tools.
4. Ready-to-copy prompt	A reusable instruction to extract data.
5. Excel template	How to turn the AI output into a useful table.
6. Limits of the manual method	When errors, friction and lack of traceability appear.
7. Signs you should automate	A checklist to know if you need custom software.
8. ROI calculator	How to estimate the monthly and yearly cost of manual work.
9. Professional solution	What a real document automation would include.
10. Use cases	Three sectors where automation usually pays off.
11. Free assessment	How to evaluate your case before building anything.

DECISION CRITERION

Not every company needs custom software from day one. But almost all of them need to know how much it costs to keep copying data by hand.

Copying data by hand looks cheap... until it repeats every week

The real cost isn't only the minutes spent, but the errors, reviews, delays and loss of control.

Many companies still receive critical information in documents that someone has to open, read, interpret and copy into Excel, ERP, CRM, accounting or an internal database.

Typical documents that create manual work

- Supplier or customer invoices.
- Delivery notes, dispatch notes, CMR, packing lists or logistics documents.
- Emails with orders, incidents, quotes or status changes.
- Contracts, annexes, receipts, forms or case files.
- Scanned PDFs, native PDFs, images and Excel sheets received from third parties.

Where the hidden cost appears

Hidden cost	Practical example
Repetitive time	Opening 200 invoices a month and copying amounts, dates and suppliers.
Copy errors	Wrong amounts, incomplete invoice numbers or incorrect dates.
Delays	Documents pending because only one person knows how to process them.
Lack of traceability	It's unclear who reviewed what, when and on what basis.
Rework	Re-checking documents because a field was missing or there was a duplicate.

Which documents you can turn into data

Document automation isn't just OCR. It's about capturing, classifying, extracting, validating and integrating.

Document	Data you can extract	Downstream use
Invoice	Supplier, VAT no., date, number, net, VAT, total, due date.	Excel, ERP, accounting, approval.
Delivery note	Customer, reference, quantities, date, delivery, issues.	Logistics control, validation against the order.
Email	Sender, subject, intent, urgency, order details.	Classification, routing, CRM, tasks.
Contract	Parties, dates, amounts, due dates, key clauses.	Alerts, follow-up, document base.
Form	Structured fields, answers, attachments, notes.	Dashboard, CRM, case file.
Logistics document	CMR, packing list, reference, origin/destination, packages.	Import/export, traceability, reporting.

BEST STARTING POINT

Start with recurring documents that have fairly similar fields and that someone currently copies by hand into Excel, ERP, CRM or accounting.

Good candidates to start with

- Recurring documents with fairly similar fields.
- Processes where information is currently copied into Excel or an ERP.
- Documents arriving by email, shared folder, portal or form.
- Tasks with human review, approval or later validation.
- Processes with frequent errors or administrative delays.

How to test it without building software

Before automating, you can check whether the AI understands your documents and which fields are worth extracting.

1. Pick 5 to 10 real, representative documents.
2. Define which fields you need to extract and in what format.
3. Upload the document to ChatGPT, Claude, Gemini or another tool that can read files.
4. Ask for structured output as a table, CSV or JSON.
5. Manually review the extracted data.
6. Copy the result into Excel and check whether it saves you time.
7. Note errors, doubtful fields and documents that aren't processed well.

WHEN THIS METHOD WORKS

It works well for low volume, internal tests, simple documents or initial validation. It's not the best option when you need traceability, integration, security or a recurring process.

SECURITY AND SENSITIVE DATA

Before uploading documents to a public AI, anonymise personal data and review its privacy policy. If you handle invoices, payslips or files with third-party data (GDPR), use fictitious documents or a private environment. The manual method is for validation, not production.

Example prompt to extract data from documents

You can copy it and adapt it to invoices, delivery notes, orders, contracts or emails.

```
Act as a document data extraction assistant for a company.
```

```
Goal:
```

```
Extract the information from the attached document and return it in a clean table.
```

```
Fields to extract:
```

- Document type
- Supplier or issuer name
- Customer or recipient name
- Document date
- Document number
- Tax base
- VAT / taxes
- Total
- Due date, if any
- Order reference, if any
- Relevant notes

```
Rules:
```

1. Do not make up data.
2. If a field is not clearly present, write "NOT DETECTED".
3. If in doubt, add "REVIEW".
4. Return a "confidence_level" column with High, Medium or Low.
5. Return the result as a table.
6. At the end, list possible errors or fields that should be reviewed manually.

Variation to get CSV

```
Return the result as CSV using a semicolon separator.
```

```
Do not add explanations before or after.
```

```
Use exactly these columns:
```

```
file;document_type;supplier;customer;date;document_number;tax_base;vat;total;due_date;reference;confidence_level;notes
```

Adapting it to other documents

- Delivery note or order: swap the fields for customer, reference, quantities, units, delivery date.
- Email or contract: ask for sender, intent, key dates, amounts and due dates.
- In all cases: keep the rules “do not make up data”, NOT DETECTED, REVIEW and confidence_level.

IMPORTANT LIMITATION

AI can get things wrong or “hallucinate” credible-looking amounts and dates. That's why the confidence_level field and human review aren't optional: never pass data into accounting without validating it.

How to move extracted data into Excel without losing control

The key is to always use the same column structure and separate data, review and status.

Column	What it's for
file	Name of the original PDF, image or email.
document_type	Invoice, delivery note, contract, order, form, etc.
supplier_issuer	Who issues the document.
customer_recipient	Who it is addressed to.
date	Main date of the document.
document_number	Invoice, delivery note, order or contract number.
total	Total amount or main value.
confidence_level	High, Medium or Low.
review_status	Pending, Reviewed, Approved, Rejected.
notes	Errors, doubts or reviewer notes.

Good practices

- Don't mix extracted data and human notes in the same column.
- Keep a status column to know which documents have been reviewed.
- Always keep the original file name.
- Use closed lists for fields like document_type or review_status.
- Don't delete failed cases: they're the most useful for designing real automation.

Mistakes to avoid

- Trusting the AI 100% without reviewing critical amounts and dates.
- Uploading documents with sensitive data without anonymising or checking privacy.
- Changing the column structure for each batch: you lose comparability.
- Not recording who reviewed each document or when.
- Deleting failures instead of keeping them as a basis for automation.

Where the manual AI method starts to break down

Manual AI is fine to start with, but it doesn't replace a robust process as volume grows.

Limit	What happens in practice
Volume	Uploading documents one by one stops paying off.
Varied formats	Each supplier or customer sends different documents.
Poor scans	OCR can fail with low quality, shadows or rotation.
Traceability	There's no clear history of extraction, review and approval.
Security	You must control what data is uploaded, where it's processed and who accesses it.
Integration	Copying to Excel isn't enough if it then has to go to an ERP, CRM or accounting.
Duplicates	AI doesn't handle repeated documents, versions or conflicts on its own.
Team	Several people processing documents need shared rules and statuses.

KEY IDEA

Manual AI is a good exploration tool. Professional automation starts when the process has to work the same way every day.

When you need a custom solution

Use this checklist to see if your process has moved from a manual test to an operational need.

- ☐ You process documents every week or every month.
- ☐ More than one person copies, reviews or approves data.
- ☐ Copy errors cause delays, complaints or rework.
- ☐ You need to export data to ERP, CRM, accounting, BI or a database.
- ☐ Documents arrive from several channels: email, folder, portal, WhatsApp or form.
- ☐ You need to detect duplicates or already-processed documents.
- ☐ You want to keep traceability of extraction, review and approval.
- ☐ Some documents have variable fields or different formats per supplier/customer.
- ☐ You want a dashboard to see pending items, errors, amounts or statuses.
- ☐ The process relies too much on one specific person.

Quick decision rule

Situation	Recommendation
Fewer than 20 documents/month and low complexity	Manual AI method + Excel.
20-100 documents/month with required review	Excel template + semi-manual flow + assessment.
100-500 documents/month	Partial automation with extraction, validation and status control.
More than 500 documents/month or several users	Custom document system integrated with internal tools.

Work out how much manual document processing costs you

You don't need a complex formula. Start by estimating time, volume and cost/hour.

Monthly manual cost = documents/month × minutes/document ÷ 60 × cost/hour

Variable	Example	Your value
Documents per month	300	
Minutes per document	5	
Company cost per hour	€25	
Monthly manual cost	$300 \times 5 \div 60 \times 25 = €625$	
Annual manual cost	$625 \times 12 = €7,500$	
Estimated saving if you automate 70%	€5,250/year	

Example: a company that processes 300 documents a month, takes 5 minutes per document and has a company cost of €25/hour spends around €625 a month on the manual process. If automation cuts that work by 70%, the potential saving is around €5,250 a year, before counting reduced errors, delays and rework.

What a custom document automation can do

It's not just about reading PDFs. It's about building a complete flow from document intake to validated data.

Stage	What the system does
Capture	Reads emails, folders, forms or portal uploads.
Classification	Detects whether the document is an invoice, delivery note, contract, order, receipt, etc.
OCR / reading	Extracts text from native PDFs, scans or images.
Extraction	Gets key fields and tables according to business rules.
Validation	Checks required fields, amounts, duplicates, formats and consistency.
Human review	Sends only doubtful or low-confidence documents for review.
Export	Sends data to Excel, ERP, CRM, accounting, BI or a database.
Dashboard	Shows pending items, errors, amounts, times and statuses.
Traceability	Records what was processed, when, by whom and with what result.

Typical architecture

- Input: email, shared folder, portal, form or API.
- Processing: OCR, AI, validation rules and duplicate detection.
- Review: a simple screen to accept, correct or reject documents.
- Output: Excel, ERP, CRM, accounting, dashboard or database.
- Control: history, logs, permissions, statuses and alerts.

KEY DIFFERENCE

OCR extracts text. Document automation turns documents into usable data, validated and integrated with your real process.

Three cases where automation usually pays off

These examples help you recognise processes where a custom solution can deliver operational ROI.

Case	Problem	Possible automation
Accounting firm	Receives invoices from many clients by email and shared folders.	Classification, data extraction, duplicate detection and accounting export.
E-commerce	Handles invoices, delivery notes and supplier documents in different formats.	Extraction of references, quantities and amounts; validation against orders.
Services company	Receives contracts, quotes, receipts and client documents.	Document portal, review statuses, alerts and case dashboard.

RECOGNISE YOURSELF IN ANY OF THESE?

If your day-to-day looks like these cases, in 30 minutes we can tell you what to automate first and how much you could save. Write to us at info@gautic.com.

Free document automation assessment

Before building anything, it's worth understanding whether your process is worth automating and where to start.

At Gautic we can help you review your current document process and decide whether to keep a manual AI-assisted method, improve your control Excel or build a custom solution.

In 30 minutes we review

- Which documents you process and how often.
- How much time your team spends copying or reviewing them.
- What errors or delays currently appear.
- Which tools you use: Excel, ERP, CRM, email, folders, portal, accounting.
- Which part can be automated first.
- Which part should keep human review.
- What kind of solution would make sense for your volume.

NEXT STEP

Request a free 30-minute assessment and we'll tell you which process to automate first, what saving to expect and whether you really need custom software. No commitment and no jargon: just clear numbers and a concrete plan.



Document automation · Integrations · AI applied to real processes

[Request a free assessment by email](#)

info@gautic.com · gautic.com · [Book your 30-min assessment](#)