

AI Invoice Extraction Test Kit

Test any AI or OCR tool on the same 10 fictional invoices, score it field by field, and decide with your own numbers instead of a vendor's.

Foxpedite does this automatically

This kit is the manual way to test invoice AI. Foxpedite puts extraction inside a checked payment flow:

- ✓ **Extraction built in:** AI document scanning (Advanced plan) reads POs, invoices and receipts, and a person reviews the fields. Put it in the Tool C column and score it on these same 10 invoices.
- ✓ **What happens after extraction:** an invoice is released for payment only when PO + receipt + invoice match (Gate 3), so a misread total or invented PO number doesn't pay itself.
- ✓ **Your data never trains public models.**
- ✓ **Matched bills sync to QuickBooks Online** (Advanced plan), so nobody re-keys what the AI read.

Try it free for 14 days: foxpedite.com/register →

Every Advanced feature, no credit card. Approval workflows, approve-from-email, QuickBooks Online sync and AI document scanning are on Advanced (\$740/mo, 20 users incl.); Regular (\$390/mo, 10 users) covers POs, budgets and 3-way match. +\$39/extra user · prices as of Oct 2026.

What's in the kit

- **10 sample invoices (PDF, in the sample_invoices folder):** fictional companies, 10 different layouts, from clean to a skewed, stamped scan.
- **Scoring sheet (.xlsx):** answer key for 13 fields × 10 invoices, columns for 3 tools, automatic accuracy formulas.
- **This guide:** the method, so your results are comparable.

Why test at all? Every vendor publishes an accuracy number, and none of them were measured on your invoices. Foxpedite's own figure (99.2% line-item accuracy) comes from Foxpedite's internal testing. This kit lets you check anyone's claim, including ours.

Fields scored

Field	Type
Vendor name critical	text
Invoice number critical	text
Invoice date critical	date
Due date	date
PO number critical	text
Bill-to company	text
Payment terms	text
Number of line items	number
First line description	text
Subtotal critical	amount
Tax critical	amount
Total / amount due critical	amount
Remit-to bank acct (last 4) critical	text

The method (30–60 minutes per tool)

- 1 Pick up to 3 tools.** For example: a general chatbot (ChatGPT, Claude, Gemini), an OCR or receipt app, and a purpose-built AP tool on a free trial. Type their names into row 6 of the Scoring tab.
FOXPEDITE: start a 14-day Foxpedite trial and run the same 10 invoices through AI document scanning as Tool C.
- 2 Same input, every time.** Upload the PDF exactly as provided, one invoice per fresh session. Don't crop, rotate or re-scan INV10. For chat tools, paste the standard prompt below. For AP tools, use their normal upload flow.
- 3 No hints, no retries.** Record the first answer. If the tool asks a question, answer only "use what's on the document." Retrying until it's right measures your patience, not the tool.
- 4 Paste values into the yellow column** for that tool. Enter dates as dates or YYYY-MM-DD; amounts with or without \$ and commas.
- 5 Time it.** Record total minutes for all 10, *including* checking and fixing. That's the number that matters for your team.
- 6 Read the Summary tab.** Look at critical-field accuracy first. A tool that's 95% accurate overall but misses the total or the bank account on 1 of 10 invoices is a payment risk, not a time-saver.

Standard prompt for chat tools

Extract these fields from the attached invoice and return JSON only: vendor_name, invoice_number, invoice_date (YYYY-MM-DD), due_date (YYYY-MM-DD), po_number, bill_to, payment_terms, line_count, first_line_desc, subtotal, tax, total_due, remit_last4. Use null if a field is not on the document. Do not guess.

The 10 invoices and the trap in each

ID	Vendor (fictional)	What it tests
INV01	Bluegill Office Goods Co.	Single page, clear table, everything labeled.
INV02	Redfern Lumber & Fastener Supply	12 lines, tax + delivery charge, dense table.
INV03	Quillmark Advisory Partners	Services: fractional hours, no sales tax, mileage line.
INV04	Tallowind Facility Services Inc.	Ship-to, bill-to AND a separate remit-to lockbox address. Tax shown as a flat amount.
INV05	Pinecrest Medical Distributors	Negative lines (discount, credit memo) and early-payment terms.
INV06	Ironbark Industrial Parts Ltd.	18 lines across 2 pages; totals only on page 2.
INV07	Marigold Restaurant Supply Co.	Landscape; qty ordered vs qty shipped vs backordered - bill is on SHIPPED.
INV08	Juniper Vale Design Studio	No table borders, long-form dates, no explicit 'Subtotal' label style.
INV09	Clearwater Valley Power & Light	Utility layout: previous balance, payment, current charges, amount due. No PO.
INV10	Old Harbor Janitorial Supply	Scanned look: skewed, grainy, low contrast, RECEIVED stamp over the totals area.

How the scoring works

- **Amounts** are correct within \$0.01. "\$1,021.84" and "1021.84" both count.
- **Dates** are compared as dates, so "12/16/26" and "2026-12-16" both count.
- **Text** ignores case, spaces, dashes, dots, commas, # and slashes. "RL-0099731" = "RL 0099731".
- **INV09 (utility)**: the correct total is the *amount due*, not current charges. It has no PO: if the tool returns null or blank, type **(none)** in its cell. A made-up PO scores 0.
- **INV07**: billed on quantity *shipped*. A tool that multiplies ordered quantities is wrong.
- Disagree with a strict match? Overwrite the score with 1 and write why in the Reviewer note. Be consistent across tools.

Reading your results

If you see...	It usually means...	What to do
High overall, misses on totals/tax	The tool reads text well but doesn't reason about the totals block	Require a human check of totals, or a 3-way match before payment
Made-up PO numbers or bank digits	Hallucination: the model filled a blank	Disqualifying for unattended use. AI should flag, not guess. In Foxpedite, an invoice has to match a real PO and receipt before it pays.
Fails only on INV10 (scan)	Weak OCR on low-quality images	Ask vendors to email PDFs; keep scans for manual entry
Fails on INV06 page 2	Only reads the first page	Check how often your real invoices run 2+ pages
Credit lines or tax read as charges	The tool doesn't separate credits, discounts and tax from line items	Score INV05 strictly; credit memos and tax lines break this in real life
Vendor name doesn't match your list	"Inc." vs "LLC", DBA names, remit-to names	Match vendors on tax ID or vendor number, not name alone

Also common in real invoices and not in this kit: handwritten notes or quantities, crooked phone photos, and the same invoice sent twice in two formats. Add a few of your own if they're common for you.

Before you upload real invoices anywhere

- ☐ Is my data used to train the vendor's or a third party's models? (Get it in writing.)
- ☐ Where is it stored, for how long, and can I delete it?
- ☐ Does the tool show its confidence, or flag fields it couldn't read?
- ☐ Can a human review and correct before anything is paid?
- ☐ Does extraction connect to a PO and a receipt, or does it just produce data someone has to re-key?

FOXPEDITE: your data never trains public models; extracted invoices are matched to the PO and receipt before payment.

Stop doing this by hand

Extraction is step one. The risk is what happens next. In Foxpedite, AI document scanning (Advanced plan) reads POs, invoices and receipts, a person reviews the result, and Gate 3 only releases an invoice for payment when PO + receipt + invoice match. Your data never trains public models. Run these same 10 invoices through the trial and score it.

Try it free for 14 days: foxpedit.com/register →

foxpedit.com/register · Trial includes every Advanced feature. Approval workflows, QuickBooks Online sync and AI document scanning are on Advanced (\$740/mo, 20 users incl.); Regular (\$390/mo) covers POs, budgets and 3-way match. +\$39/extra user

Notes: All companies, people, addresses, account numbers and amounts in the sample invoices are fictional and created for this kit; any resemblance to real businesses is coincidental. The 99.2% figure is from Foxpedite's internal testing and is not independently verified. Tool names are examples, not endorsements.