FLOWGRAMMER FIELD WORKBOOK

Invoice Processing Test Pack

Five fictional cases to test extraction, duplicate controls, human review, and safe draft creation.

The rule: No invoice reaches an accounting destination until a person approves it. A retry must reuse the same idempotency key.

VERSION	TEST DATE	OWNER
0.1.0	2026-09-08	Flowgrammer

BUILD THE SYSTEM. TEST THE FAILURE PATHS. KEEP THE FINAL DECISION HUMAN.

01 / START HERE

What this pack proves

A small acceptance suite for the steps between invoice intake and an accounting draft.

Use the same five cases against every extractor or workflow. Record the output. Compare it with the expected result. Keep each score separate so a strong extraction result cannot hide an unsafe workflow.

Run the baseline

1. Open each file in fixtures/ and its matching file in expected/.
2. Run your extractor or map its output into schema/invoice-output.schema.json.
3. Run python3 run_tests.py to confirm the reference behaviour.
4. Record your own result in the scorecard on page 9.
5. Investigate every failed safety check before connecting a live destination.

Scope: These fixtures are labeled text. They test workflow logic and field mapping. They do not prove OCR quality on scans or PDFs.

What good looks like

CONTROL	PASS CONDITION
Extraction	Every expected field matches the fixture.
Duplicate handling	A repeated source or vendor + invoice number stops before review.
Human review	Every non-duplicate waits for an explicit decision.
Destination safety	A lost response can retry without creating another draft.
Audit trail	The result records status, reasons, attempts, writes, and test date.

02 / TEST MAP

The five cases

Each case isolates one condition that can break an invoice workflow.

CASE	FIXTURE	EXPECTED RESULT	DESTINATION
1. Clean	INV-TP-2401 CAD 452.00	Wait for review, then draft after approval	1 attempt / 1 write
2. Duplicate	Replay of INV-TP-2401	Stop before review	0 attempts / 0 writes
3. Missing number	Ashbridge Print Works CAD 257.64	Flag missing_invoice_no; reject in reference run	0 / 0
4. Conflicting total	INV-TP-2404 printed CAD 185.00	Flag conflicting_total; reject in reference run	0 / 0
5. Lost response	INV-TP-2405 CAD 164.98	Approve; retry with same idempotency key	2 attempts / 1 write

Your test environment

ITEM	RECORD
Extractor and version	_____
Workflow and version	_____
Destination or sandbox	_____
Reviewer	_____
Date tested	_____

Protect live systems: Use a sandbox or fake destination first. The pack never posts, pays, refunds, or changes supplier records.

03 / EXTRACTION

Field checklist

Map the extracted object to the included schema before judging workflow behaviour.

FIELD	REQUIRED HERE	CHECK
vendor	Yes	[ ]
vendor_address	No	[ ]
tax_registration	No	[ ]
invoice_no	Yes	[ ]
date	Yes, YYYY-MM-DD	[ ]
due_date	No	[ ]
po_number	No	[ ]
currency	Yes, three letters	[ ]
payment_terms	No	[ ]
bill_to	No	[ ]
bill_to_address	No	[ ]
line_items	Yes; may be empty	[ ]
subtotal	Yes	[ ]
tax_label	No	[ ]
tax	Yes	[ ]
total	Yes	[ ]

Money tolerance: 0.011. This allows ordinary decimal rounding while still catching the conflicting-total fixture.

04 / CASES 1 AND 2

Clean invoice and duplicate

Prove the happy path without weakening the duplicate boundary.

Case 1: clean invoice

EXPECTED FIELD	VALUE
Vendor	Ridgemont Industrial Supply Co
Invoice	INV-TP-2401
PO	PO-CQ-188
Subtotal / tax / total	CAD 400.00 / 52.00 / 452.00
Without reviewer	pending_review; 0 destination attempts
After explicit approval	drafted; 1 attempt; 1 write

Case 2: duplicate

- Replay the byte-identical INV-TP-2401 fixture.
- Also test the same vendor and invoice number with changed file bytes.
- Expect duplicate status before any reviewer or destination call.

OBSERVATION	PASS	NOTES
Source hash duplicate stopped	[ ]	_____
Vendor + invoice number duplicate stopped	[ ]	_____
Reviewer invocations stayed at zero	[ ]	_____
Destination attempts stayed at zero	[ ]	_____

05 / CASES 3 AND 4

Exceptions that need a person

The workflow can detect a problem. A person owns the decision to proceed.

CASE	SIGNAL	REFERENCE RESULT
Missing invoice number	invoice_no is null	pending_review without a reviewer; rejected by the simulated test decision
Conflicting total	210.00 + 10.50 does not equal printed 185.00	pending_review without a reviewer; rejected by the simulated test decision

Human means human: The included simulated reviewer makes the tests repeatable. It is not permission for a production workflow to decide on its own.

Review record

QUESTION	ANSWER
Which source did you compare?	_____
What caused the exception?	_____
Approve, edit, or reject?	_____
Who made the decision?	_____
What evidence supports it?	_____

06 / CASE 5

A timeout after the write

Test the failure that creates duplicates when retries are careless.

The fake destination accepts INV-TP-2405 and creates one draft. Its response is lost. The workflow retries with the same idempotency key.

STEP	EXPECTED
1. Reviewer approves	One explicit approval is recorded.
2. First request	The destination writes the draft, then the response times out.
3. Retry	The workflow reuses the original idempotency key.
4. Final count	Two attempts, one stored draft.

Evidence to retain

EVIDENCE	VALUE
Idempotency key on attempt 1	_____
Idempotency key on attempt 2	_____
Destination attempts	_____
Destination writes	_____
Final status	_____

Pass condition: Both requests carry the same key. The destination records one draft. A different key or a second draft is a failed safety test.

07 / FAILURE PATHS

Control checklist

Review the edges before a live inbox, OCR service, or accounting platform is connected.

FAILURE	EXPECTED CONTROL	PASS
No reviewer available	Stay pending_review; destination remains untouched	[ ]
Duplicate source hash	Stop before review	[ ]
Duplicate vendor + number	Stop before review even if bytes changed	[ ]
Missing invoice number	Require review; reference decision rejects	[ ]
Arithmetic conflict	Require review; reference decision rejects	[ ]
Reviewer rejects	Record rejected; zero destination attempts	[ ]
Write response lost	Retry with same key; one total write	[ ]
Malformed adapter JSON	Reject wrong Boolean, array, or integer types	[ ]

Extensions for a later test set

- Unreadable or rotated scan
- Low-confidence vendor or invoice number
- Credit note mistaken for an invoice
- Foreign currency and tax treatment
- Supplier bank-detail change
- Multi-page invoice with a missing page

08 / SCORECARD

Keep five scores separate

A workflow is only ready when extraction and safety pass together.

DIMENSION	HOW TO SCORE	RESULT
Extraction accuracy	Matched expected fields / expected fields x 100	_____ %
Workflow safety	Legal state transitions and no illegal writes	_____ %
Human review	Non-duplicates wait; exceptions record a decision	_____ %
Duplicate handling	Both duplicate routes stop before review	_____ %
Destination reliability	Timeout produces 2 attempts and 1 write	_____ %

Do not average away a safety failure: A 100% extraction score does not compensate for an automatic write, duplicate draft, or missing review record.

Decision

CHECK	DECISION
[ ]	Ready for a sandbox integration
[ ]	Needs field-mapping work
[ ]	Needs workflow-control work
[ ]	Needs human-review design
[ ]	Stop: unsafe destination behaviour

Owner: _____ Review date: _____ Next retest: _____

09 / NEXT STEP

Turn a passing test into an owned system

Use the pack to identify the exact boundary between extraction, review, and accounting.

Keep the workflow small

1. Choose one intake channel.
2. Choose one extraction method.
3. Map to one documented invoice schema.
4. Route every non-duplicate through a human decision.
5. Create a draft only after approval.
6. Log the result and measure exceptions.

When a template stops being enough

You need an owned workflow when invoices arrive through several channels, a live ERP write is involved, supplier bank details are stored, or one duplicate can create material cost.

Flowgrammer: We build documented invoice intake-to-accounting systems with human gates, duplicate controls, monitoring, and handover. Start with an AI Success Audit when the workflow or priority is still unclear.

Files in the technical pack

- Five fictional fixtures and expected JSON results
- Invoice and adapter JSON schemas
- Reference processor and local test runner
- Scoring rubric and adapter field map
- Licence proposal, notice, and release-readiness notes

Related guide: flowgrammer.ca/insights/document-processing-automation

10 / METHOD

What this workbook does and does not claim

A transparent baseline for comparison, not a production benchmark.

Method

The reference runner extracts fixed labeled text, validates required fields and arithmetic, blocks duplicates, waits for review, and simulates an ambiguous destination timeout. The fixtures are fictional and contain no personal or client data.

Limits

- No OCR engine or external model was called.
- No Azure, OpenAI, Anthropic, Slack, ERP, or Supabase service was called.
- No scan, photograph, handwriting, or low-confidence case is included.
- A perfect result on five fixtures is not evidence of production accuracy.
- MIT is a proposed release licence until the owner approves a public release.

References

- Flowgrammer. Document Processing Automation. Accessed September 8, 2026. flowgrammer.ca/insights/document-processing-automation
- Microsoft Learn. AI document processing reference architecture. Accessed September 8, 2026.
- n8n. Process invoices securely with OpenAI, Slack, and ERP webhooks, template 17 806. Accessed September 8, 2026. Architecture reference only; its workflow JSON is not included.

Licence and reuse

This local technical package proposes the MIT License. Public release still requires owner approval. See LICENSE and NOTICE.md in the package.

Build the system. Test the failure paths. Keep the final decision human.