

Invoice Validation Agent

From month-end scramble to audit-ready close.



Live Production Payload

STATUS: VERIFIED

VENDOR

Contractor Services

CONTRACTOR

Robert J. Smith

HOURS LOGGED

120 hrs

CONTRACT RATE

\$145 / hr

< 1 min

Invoice arrival to final audit-ready decision



Month-end runs on **spreadsheets**.

One invoice. Six manual handoffs.

01	 A/P Team	Rekey names, hours, rates & dates
02	 Spreadsheet	Track individual invoice details line-by-line
03	 Broadcom Clarity	Find contractor & compare records manually
04	 Vendor Comms	Chase down discrepancies over email chains
05	 Spreadsheet	Update notes, manual evidence & audit markers
06	 Final Approval	Manually approve and hand off to payment run

10–15 min

Wasted per single contractor invoice

Upwards of 25%

Of total senior analyst time drained

Audit Vulnerability

Evidence trapped in ad-hoc spreadsheet cell notes

Matching is the hard part.



INCOMING VENDOR INVOICE

Contractor	Robert J. Smith
------------	------------------------

Logged Hours	120 hours
--------------	------------------

Billed Rate	\$145 / hour
-------------	---------------------

Billing Period	June 1 – 30
----------------	--------------------

BROADCOM CLARITY SYSTEM

Candidate 1	Bob Smith
-------------	------------------

Exact Match	Robert Smith (RES-4471)
-------------	--------------------------------

94% CONF.

Candidate 3	R. J. Smith
-------------	--------------------

Timesheet Status	TS-88214 · Approved
------------------	----------------------------

Tune on real history

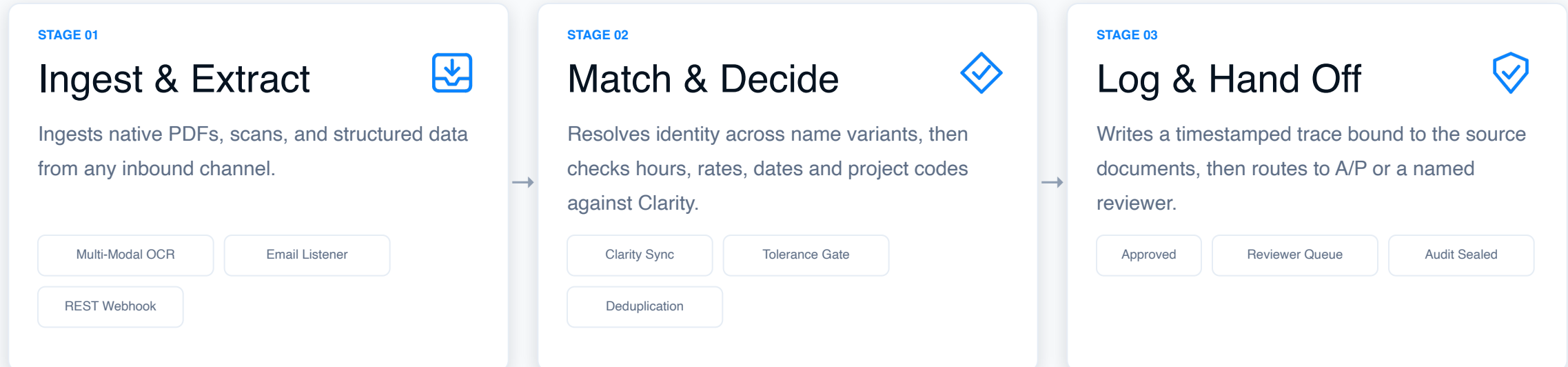
Fuzzy matching needs production invoice history to calibrate threshold tolerances accurately.

Test messy formats

Extraction was already solved; identity resolution requires testing messy edge cases, not synthetic cleans.



An agent that closes the loop.



< 1 minute from arrival to audit-ready close.



Keep **humans** in the loop.

AUTOMATED GATING SEQUENCE

1 Match confidence sufficient?

2 Rate & hours within tolerance?

Automated close execution

Start conservative, widen with evidence

Route exceptions to named reviewers

Trust builds one clean quarter at a time

Business sets the boundaries. Agent automates the obvious.
Humans own the exceptions.



Hours back, errors out.

PROCESSING SPEED TODAY

10–15 min

Per invoice manual verification, spreadsheet double-entry, and email chasing.



AGENTIC SPEED

< 1 min

End-to-end ingestion, Clarity matching, rule validation, and audit seal.

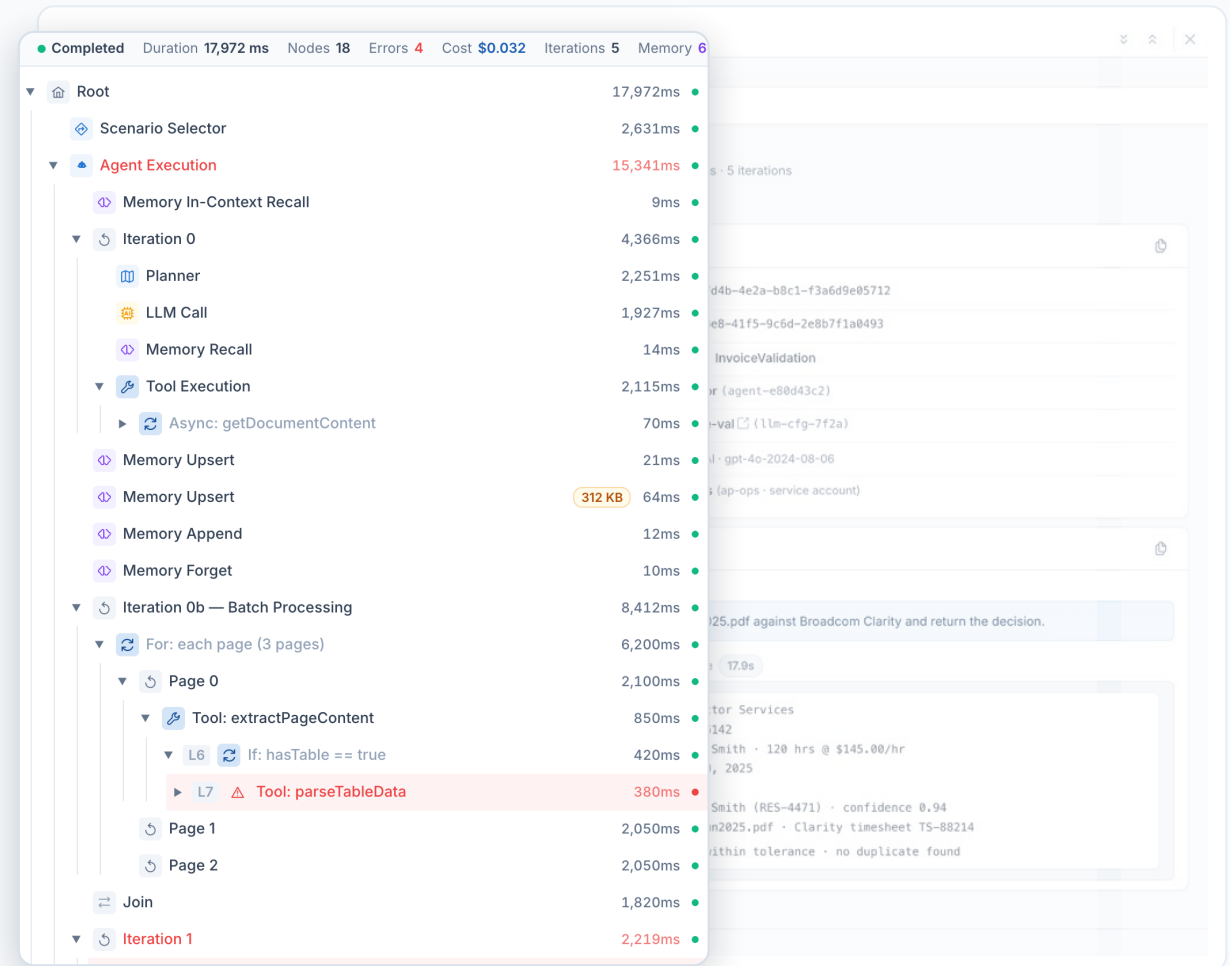
1,000s

Invoices validated in batched runs each month-end close.



Timestamped decision logs.

Every prompt, tool call and result is serialized into an immutable, queryable trail. This one invoice: 18 nodes, 5 iterations, every timing on the record.



Probe · full run trace — every node timed



Reasoning & source, [linked](#).

The agent's reasoning is captured verbatim and bound to its source — here it reads the vendor's PDF and emits structured, checkable fields.

The screenshot displays the 'Probe' interface, which provides a detailed view of an agent's execution. On the left, a tree view shows the execution flow: Root (17,972ms) -> Scenario Selector (2,631ms) -> Agent Execution (15,341ms) -> Iteration 0 (4,386ms) -> Planner (2,251ms) -> LLM Call (1,927ms) -> Tool Execution (2,115ms) -> Async: getDocumentContent (70ms) -> Iteration 0b - Batch Processing (8,412ms) -> For: each page (3 pages) (6,200ms) -> Join (1,820ms) -> Iteration 1 (2,219ms) -> LLM Call (1,102ms) -> Tool Execution (1,117ms) -> Iteration 2 (3,582ms) -> LLM Call (2,450ms) -> Tool Execution (1,132ms) -> Iteration 3 (2,672ms) -> LLM Call (1,600ms) -> Iteration 4 (2,497ms) -> LLM Call (1,350ms).

The right panel shows a detailed view of the selected LLM Call (GPT-4o, 1,927 ms, status: 200). It includes a 'RESPONSE' section with a 'Response Code' of 200 and an 'LLM THOUGHT' section. The thought text reads: 'The user wants to extract invoice details from a PDF. I need to parse the document content and identify key invoice fields: vendor name, invoice number, dates, line items with descriptions and amounts, subtotal, tax, and total. The document appears to be from Cox Communications for internet service. I will structure the output as a JSON object with clearly labeled fields.'

Below the thought, the 'DATA' section shows a JSON object:

```
{
  "vendor": "Cox Communications",
  "invoice_number": "INV-2025-03847",
  "date": "2025-10-15",
  "due_date": "2025-11-14",
  "line_items": [
    {
      "description": "Internet Preferred
250",
      "amount": 89.99
    }
  ],
  "subtotal": 89.99,
  "tax": 7.20,
}
```

Probe · LLM reasoning + structured output



Exceptions route to a human.

When a step fails or confidence drops, the run routes to a named reviewer with clear rationale, then retries and completes — all visible in the trace.

The screenshot displays the 'Probe' interface, which provides a detailed trace of an LLM execution. The top bar indicates the run is 'Completed' with a duration of 17,972 ms, 18 nodes, 0 errors, a cost of \$0.032, and 5 iterations. The left sidebar shows a tree view of the execution steps, including 'Root', 'Scenario Selector', 'Agent Execution', and several iterations of 'Planner', 'LLM Call', and 'Tool Execution'. The main panel on the right shows the 'HITL REQUESTS' section, which contains a single request labeled '1 Approval requested Iteration 0' with a green 'Approved' status. Below this, the 'CONVERSATION' section shows the user query 'Validate Invoice-VCS-Jun2025.pdf against Broadcom Clarity and return the decision.' and the agent response from 'InvoiceValidator' with a 17.9s duration. The agent response includes details about the invoice, contractor, and a final decision of 'APPROVED'.

Probe

Completed Duration 17,972 ms Nodes 18 Errors 0 Cost \$0.032 Iterations 5

Filter nodes... Errors Cost (\$) Join Tools

Root 17,972ms

- Scenario Selector 2,631ms
- Agent Execution 15,341ms
 - Iteration 0 4,386ms
 - Planner 2,251ms
 - LLM Call 1,927ms
 - Tool Execution 2,115ms
 - Async: getDocumentContent 70ms
 - Iteration 0b — Batch Processing 8,412ms
 - For: each page (3 pages) 6,200ms
 - Join 1,820ms
 - Iteration 1 2,219ms
 - LLM Call 1,102ms
 - Tool Execution 1,117ms
 - Iteration 2 3,582ms
 - LLM Call 2,450ms
 - Tool Execution 1,132ms
 - Iteration 3 2,672ms
 - LLM Call 1,600ms
 - Iteration 4 2,497ms
 - LLM Call 1,350ms

HITL REQUESTS 1 request

1 Approval requested Iteration 0 **Approved**

Agent requested human approval before executing document extraction on the uploaded invoice.

LLM Config gpt-4o-invoice-val {llm-cfg-7f2a}

LLM Model GPT-4o OpenAI · gpt-4o-2024-08-06

User A/P Operations (ap-ops · service account)

CONVERSATION

A/P Operations User query

Validate Invoice-VCS-Jun2025.pdf against Broadcom Clarity and return the decision.

InvoiceValidator Agent response 17.9s

Vendor: Vendor Contractor Services
Invoice #: INV-2025-06142
Contractor: Robert J. Smith · 120 hrs @ \$145.00/hr
Period: Jun 1 – Jun 30, 2025
Total: \$17,400.00
Clarity match: Robert Smith (RES-4471) · confidence 0.94
Source: Invoice-VCS-Jun2025.pdf · Clarity timesheet TS-88214
Decision: APPROVED · within tolerance · no duplicate found

F8 Next error Shift+F8 Prev error Ctrl+Click Open link in new tab Esc Close probe

Probe · HITL approval, failed-step retry



Answered from **memory**.

Human-in-the-loop resolutions become memory. On an identical future query the agent recalls the saved preference instead of asking again.

The screenshot displays the 'Probe' interface with a tree view on the left and a detailed view on the right. The tree view shows a hierarchy of operations: Root (17,972ms), Scenario Selector (2,631ms), Agent Execution (15,341ms), Iteration 0 (4,386ms), Planner (2,251ms), LLM Call (1,927ms), Tool Execution (2,115ms), Async: getDocumentContent (70ms), Iteration 0b — Batch Processing (8,412ms), For: each page (3 pages) (6,200ms), Join (1,820ms), Iteration 1 (2,219ms), LLM Call (1,102ms), Tool Execution (1,117ms), Iteration 2 (3,582ms), LLM Call (2,450ms), Tool Execution (1,132ms), Iteration 3 (2,672ms), LLM Call (1,600ms), Iteration 4 (2,497ms), and LLM Call (1,350ms). The detailed view on the right shows the 'Memory Recall' operation, which is successful and took 14ms. It reads the user's saved reporting preferences to shape the report. The operation is labeled 'memoryRecall' and reads across the user's memory + shared memory. The result is '1 item matched'. The filter used is 'reporting_pref' with a filter key of '{ "type": "reporting_pref" }'. A note at the bottom states: 'Recall returns contents only — the probe shows the match count and filter, not the returned values.'

Probe

Completed Duration 17,972 ms Nodes 18 Errors 0 Cost \$0.032 Iterations 5

Filter nodes...

Errors Cost (\$) Join

Root 17,972ms

Scenario Selector 2,631ms

Agent Execution 15,341ms

Iteration 0 4,386ms

Planner 2,251ms

LLM Call 1,927ms

Tool Execution 2,115ms

Async: getDocumentContent 70ms

Iteration 0b — Batch Processing 8,412ms

For: each page (3 pages) 6,200ms

Join 1,820ms

Iteration 1 2,219ms

LLM Call 1,102ms

Tool Execution 1,117ms

Iteration 2 3,582ms

LLM Call 2,450ms

Tool Execution 1,132ms

Iteration 3 2,672ms

LLM Call 1,600ms

Iteration 4 2,497ms

LLM Call 1,350ms

Memory Recall

Read the user's saved reporting preferences to shape the report.

✓ Success 14ms Read

OPERATION

memoryRecall

Reads across this user's memory + shared memory

Result: 1 item matched

FILTER

reporting_pref

FILTER KEY

{ "type": "reporting_pref" }

Recall returns contents only — the probe shows the match count and filter, not the returned values.

F8 Next error Shift+F8 Prev error Ctrl+Click Open link in new tab Esc Close probe

Probe · memory recall — 1 item matched



What we would tell the next team.

01

Matching is the hard part

Extraction was solved; identity resolution was not

Same vendor, three spellings, plan for it

Fuzzy matching needs real invoice history to tune

Test on the messy formats, not the clean ones

02

Keep humans in the loop

Tolerance thresholds must be finance-owned

Start conservative, widen with evidence

Route exceptions to named reviewers

Trust builds one clean quarter at a time

03

Audit trail wins the room

Reasoning logs turned the security review around

Auditors cared more about traceability than accuracy

Deployment choice mattered: documents never moved

Integrations, not models, drove the timeline



Beyond **invoice validation**.

Invoice validation was the beachhead; working with multiple teams, deployed in VPC.

01 / ADVISORY

Advisor
Enablement

FA Assistant

02 / CORE OPS

Operations

Invoice Assistant
Status Center

03 / GOVERNANCE

Data
Governance

LineAlage

04 / DEVOPS

Engineering
Productivity

Engineering Assistant

05 / LEGACY

Mainframe
Modernization

COBOL Intelligence

Live Demo

Watch the agent ingest, match against Clarity in real time, execute threshold logic, and seal the audit log in < 1 minute.



Bring us **a workflow**.

Take back a working agent.

-
- 01 One call from you to define the workflow
 - 02 One day from us to build and wire the agent
 - 03 Your sample data; no cost, no procurement cycle
-



dataworkz.com/regou