Evaluating Regulatory Workflow Automation Vendors (US edition)

A due-diligence checklist for government agencies and regulated teams choosing an automation vendor

Direct answer: how to evaluate a regulatory workflow automation vendor

Evaluate vendors on five things: whether humans keep decision authority (approval checkpoints, not full autonomy), whether every automated step is logged in an audit trail, where your data is processed and whether the vendor's AI trains on it, whether the system works inside your existing environment (e.g. working directly inside your Microsoft 365 or Google Workspace tenant), and whether they will prove value in a scoped pilot against your own baseline metrics before you commit. A vendor who resists any of those five is the wrong vendor for regulated work.

Why regulated buyers need a different checklist

In a regulated environment, the workflow itself is often prescribed: who may approve, what must be recorded, how long records are kept, and who can access what. Generic automation pitches optimise for speed and headcount; a regulatory buyer has to optimise for **speed without changing who is accountable**. That means the automation must wrap around your approval and records obligations — public-records laws, HIPAA (where health data is involved), SOX (where financial controls apply), and your agency or state records-retention schedules — not route around them.

> **The core principle**
>
> Automate the preparation, checking, routing, and tracking around a decision. Keep the decision itself with the authorised person, and log everything.

The 10-point vendor evaluation checklist

#	Requirement	What to ask the vendor	Evidence to demand
1	Human decision authority	Which actions require human approval? Can we define our own approval checkpoints per workflow?	A live demo of an approval gate blocking an action until sign-off
2	Audit trail	Is every automated step logged — what was done, when, and under whose approval?	An export of the actual audit log from a test workflow
3	No training on your data	Is our data ever used to train your models or anyone else's?	The contractual clause saying so, not a verbal assurance
4	Data residency & processing	Which cloud regions process and store our data? Can processing be pinned to US regions?	The list of subprocessors and processing regions in writing
5	Works in your environment	Does it operate inside our existing tenant and tools, or does our data get copied into yours?	An architecture diagram showing where data lives at each step
6	Records & exit	Can we export every record, log, and document in open formats if we leave?	A test export during the pilot, not a promise
7	Access control	How are roles and permissions enforced? Who at the vendor can see our data?	Their access-control model and internal-access policy in writing
8	Failure behaviour	What happens when the AI is uncertain or wrong? Does it stop and escalate, or push through?	A demo of an exception/escalation path on a malformed input
9	Scoped pilot	Will you prove value on one workflow, against our measured baseline, before a broader commitment?	A written pilot plan with baseline metrics and review criteria
10	Honest claims	Ask for the evidence behind any ROI or adoption statistic in their pitch.	Named methodology or a withdrawal of the claim — both are informative

Red flags that should end the conversation

- **"Fully autonomous" pitches for regulated work.** If nobody signs off, your accountability chain is broken by design.
- **Universal ROI percentages.** "Clients typically see 300% ROI" is a marketing number, not a measurement. Real ROI depends on your workflow and is measured against your baseline.
- **No inspectable audit log.** If you cannot export a record of what the system did, you cannot answer an FOI request, an auditor, or a minister.
- **Your data leaves your tenant "for processing"** with no clear account of where it goes, or terms that permit training on it.
- **Testimonial-only evidence.** Ask to speak to a reference customer with a comparable workflow, or weight the evidence accordingly.

Scoring and running the evaluation

1. **Map the approval workflow first.** Before any vendor call, write down the decision points, who is authorised to make each one, and what must be recorded. This is your requirements document — vendors respond to it instead of pitching around it.
2. **Score every vendor against the 10 points** with the evidence column as the standard — a requirement without evidence scores zero, however good the demo felt.
3. **Run a scoped pilot with the leading vendor.** One workflow, a measured baseline (volume, handling time, cycle time, error rate), human sign-off on every consequential action, and a review at the end against those numbers.

> **Procurement tip**
>
> Put the checklist in your RFQ verbatim and require written responses. The quality of a vendor's written answers on audit trails, data handling, and failure behaviour predicts the quality of the product.

How SG1 answers this checklist

We publish our own answers because we expect to be evaluated the same way: consequential actions go through human approval checkpoints; every AI step is logged with what it did, when, and under whose approval; models run on Microsoft Azure and are never trained on your data; the system works inside your Microsoft 365 environment rather than copying records out; and we start with a scoped pilot measured against your own baseline.

We deliberately publish no universal ROI percentage — point 10 applies to us too.