



FIELD NOTES · 7 MIN READ

# Lessons from the Front Lines: Building Audit Agents That Hold Up

Field notes from building AI audit and reconciliation agents for insurance and finance back offices — what the historical data reveals, why always-on beats quarterly, where the token costs hide, and why we build inside the client's environment.

By **StrataEdge AI** · Published June 9, 2026 · [strataedge.ai](https://strataedge.ai)

Over the past two years we have designed and shipped AI audit and reconciliation agents for insurance and finance operations — including a premium and claims audit agent for a global specialty insurance group. The system runs in production: 99% automated accuracy, with human validation taking the result to roughly 100%, and quarterly audit time cut by 75%.

These are the lessons we keep relearning on every engagement. None of them are about model selection. All of them are about whether the system holds up when the numbers matter.

## 1. The first thing your agent finds is errors in the past

Every audit agent gets built and tuned against historical audits — work that was completed, reviewed, and accepted, sometimes years ago. And every time, the build surfaces significant errors in those past audits.

This changes the conversation. The project stops being a cost-reduction exercise and becomes a risk discovery: the errors were always there, hiding in work everyone trusted. Plan for this. Decide before the build starts who gets told when the agent finds a material discrepancy in a closed quarter, because it will.

## 2. Audit wants to be always-on

A quarterly audit is a sampling exercise shaped by what a team of humans can read in a few weeks. An agent doesn't sample — it reads everything, every day.

That turns audit from a periodic detection event into a continuous property of the system: errors get caught when they enter, not months later when they have compounded across statements. The same agent that cuts the quarterly audit to a fraction of its former time also makes it progressively less of an event, because by quarter's end there is little left to find. Error avoidance replaces error archaeology.

### 3. Accuracy is a systems property, not a model property

---

The model is the easy part. Hitting accuracy your CFO will sign off on requires a deep understanding of the operational areas the audit covers — how commissions actually get calculated, what a bordereau looks like when an MGA's spreadsheet exports break, which discrepancies are policy exceptions and which are errors.

The architecture that delivers it is unglamorous and essential:

- A confidence floor below which the agent does not act alone
- Human review routed to exactly the items that need judgment, with an interface built for deciding rather than reading
- A decision trace on every conclusion, so any number can be walked back to its sources
- A fallback path — rules plus a manual queue — for the day something upstream changes shape

That is how 99% automated becomes ~100% delivered: the system knows what it doesn't know, and says so.

### 4. Token costs are real money, and the stack decides them

---

LLM costs can get out of hand quietly — especially when latency stretches a workflow across hours or days and the system compensates with retries and re-reads. We have reviewed systems carrying three times more model cost than the work required, purely because of how the calls were structured.

Streamlining the call pattern — what gets read once and cached, what gets summarized before it travels, which steps need a frontier model and which need none at all — routinely improves both performance and unit economics. Token efficiency is an engineering discipline, and it determines whether the business case survives contact with the invoice.

### 5. Build inside the client's environment

---

Audit work runs on the most sensitive material a company has: contracts, financials, claims, correspondence. Shipping that to an external service is how promising projects die in security review.

We build inside the client's own cloud environment. Documents and financial data never leave the perimeter; security and compliance teams review infrastructure they already govern; and the system inherits the controls — access, retention, residency — the organization has already paid to establish. It is the difference between asking for an exception and asking for a login.

### 6. The people part is the quiet win

---

The measurable results get reported to the board — a quarter of the audit time, twice the throughput. The result nobody puts in the deck: morale. The agent absorbs the tedious document comparison work, and the operations team spends its time on the judgment calls that were always the actual job. Adoption follows, because the system makes the work better rather than threatening it.

## What we'd tell a finance leader

---

If you are watching AI get funded across the company and asking where your value is: the audit and reconciliation layer is the highest-yield place we know. The work is document-heavy, rule-bound, and measurable — exactly what agents do well — and the payoff arrives as accuracy and control, not just cost.

Start with one workflow where the numbers have to be right. Insist on the controls being designed in from day one. Expect the historical data to surprise you. And measure the result the way an auditor would — because now you can.