

Know what your AI is doing. And what it is worth.

AI agents are in production, and the questions have changed. Is it working? When it gets something wrong, who is accountable, the agent or the person who approved it? And is it paying back what it costs? **Evolvable.ai answers all three:** it traces every run end to end, attributes every decision to an agent version or a named person, and measures the value against baselines you set.

ONE RUN, TRACED AND ATTRIBUTED



THE VALUE OF THIS RUN

MANUAL BASELINE

45 min

HUMAN TIME SPENT

14 min

approval + review

AI COST

\$0.04

NET TIME SAVED

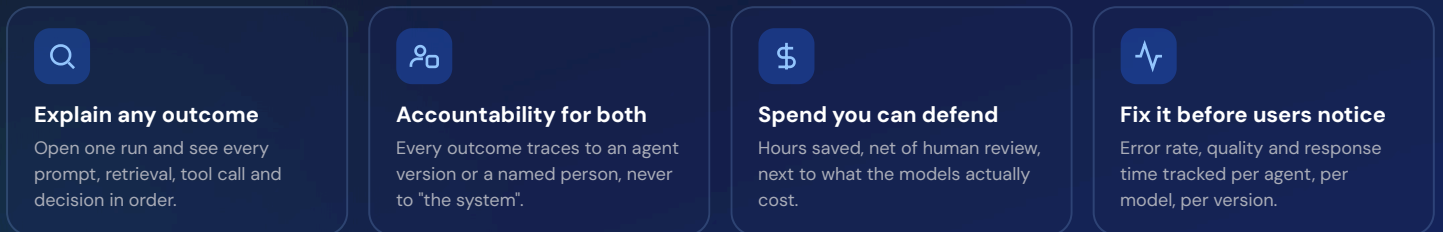
31 min

Rolls up to

Finance Ops · Refunds

Illustrative figures

WHAT YOU GET



BUILT FOR

CIO & Chief AI Officer

One view of every agent, its quality and its cost.

CFO & finance

A return on AI that holds up to scrutiny.

COO & business owners

Know which processes AI improves, and by how much.

AI & platform teams

Debug, tune and compare agents with real evidence.

QUESTIONS YOU CAN ANSWER IN ONE CLICK

Which agent costs the most per resolved case?

Did version 15 make the answers worse?

Who approves slowest, and is the SLA holding?

How often does the AI judge escalate to a person?

What did Finance Ops save last quarter?

Which model is worth its price?

4

quality KPIs on every agent and model, tracked over time

3

levels of drill-down: fleet, agent, single run

2

kinds of accountability: the agent version and the person who decided

1

ROI model, with your own baselines and hourly rates

Trace. Monitor. Account. *Measure.*

Conversation trace

Finance Assistant · v14

- Prompt** "Refund the duplicate charge on #4471"
- Retrieval** Refund policy v3 · 2 chunks
- Tool call** get_transactions(#4471)
- Approval** refund_payment · approved in 11 min
- Response** Refund issued, customer notified

12.4k tokens · \$0.04 · 18 s · 0 guardrail hits

01 TRACE

Why did the agent do that?

OUTCOME *Any outcome explained in minutes, not days.*

- ◆ **End-to-end traces:** prompts, retrievals, tool calls, sub-agent hand-offs and approvals, with the agent version that ran.
- ◆ **Every conversation and workflow run** searchable by agent, user, channel and status.
- ◆ **Raw logs** of every model call and tool call, for the engineer who needs the detail.

Agents breakdown

Last 30 days

AGENT	RESP.	ERRORS	GUARDRAIL	EVAL PASS	TREND
Finance Assistant	2.1 s	0.2%	1.1%	98%	
SOC Triage Agent	9.8 s	0.0%	0.0%	95%	
Customer Support	23.5 s	4.8%	2.3%	81%	

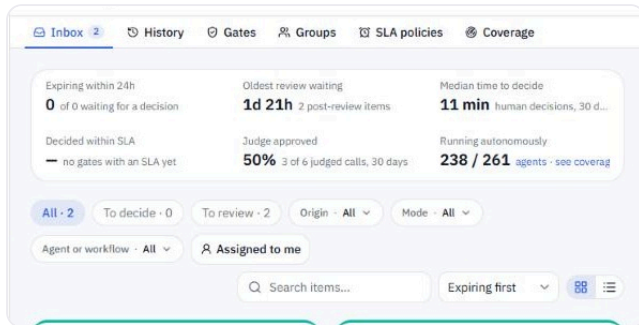
Customer Support · error rate up since v9 · open traces

02 MONITOR

Are agents fast, accurate and reliable?

OUTCOME *Problems found before users report them.*

- ◆ **Quality KPIs:** response time, error rate, guardrail violation rate and eval pass rate, per agent and model.
- ◆ **Usage & tokens:** conversations, messages, tool calls and token use, by model.
- ◆ **Live supervision:** flagged conversations routed to supervisors, with escalation.



03 ACCOUNT

Who decided: the agent or a person?

OUTCOME *A named owner behind every outcome.*

- ◆ **Every decision attributed** to an agent version, or to an approver, their group and the channel used.
- ◆ **Human performance:** median time to decide, decisions within SLA, AI-judge approval rate, reviews waiting.
- ◆ **Oversight that scales:** veto windows and AI judges cut approval queues without removing people.

ROI by business unit

Last 90 days

HOURS SAVED	COST SAVED	AI SPEND	RETURN
1,240 h	\$62,000	\$3,100	20x
net of 86 h human review	at your hourly rates	tokens × model pricing	
Finance operations			\$27,500
Security operations			\$21,000
Customer service			\$13,500

Baselines and hourly rates set by you, per task, in the configurator

04 MEASURE

Is AI paying back what it costs?

OUTCOME *A return on AI you can defend to the CFO.*

- ◆ **Your baselines:** set the manual time and hourly rate per task; savings are counted net of human review.
- ◆ **Real AI spend:** token counts priced with your own model rates, for every agent and workflow.
- ◆ **Roll-ups by business unit:** a tag tree from department down to task, with CSV export.

Visuals are illustrative, except Account (product screenshot). Figures are examples.

SOVEREIGN BY DESIGN

- Runs on your cloud, private cloud, on-premises or air-gapped. Traces and usage data stay in your estate.
- Flat annual licence, unlimited users. Bring your own models and price them at your own rates.

OBSERVES EVERY MODEL YOU RUN

OpenAI

Azure OpenAI

Anthropic

Google Gemini

Ollama & open-weight

*See what your AI is worth.
In 45 minutes.*

We connect one live agent, agree a baseline with you, and walk you through the trace, the accountability trail and the ROI.

[BOOK A DEMO](#)

info@evolvable.ai

www.evolvable.ai