medevolve

AI Can’t Fix RCM Issues You’ve Never Measured

A Revenue Cycle Leader’s Guide to Vetting AI Vendors on Outcomes, Not Task Completion

The AI Promise vs. The Measurement Problem

Revenue cycle leaders are being pitched AI as the fix for denials, staffing gaps, and administrative burden — from every direction, in every sales cycle, all at once. But most AI vendors are optimizing for task completion, not financial outcomes. And if an organization has never measured whether its own work actually drives payment, no vendor — however sophisticated the model — can prove, or deliver, real ROI on top of it.

95% Fail to Deliver

Roughly 95% of generative AI pilots fail to produce measurable P&L impact enterprise-wide, according to MIT's 2025 "GenAI Divide" research — and industry commentary cites a similar 95% failure rate for healthcare AI projects specifically.

ROI Goes Unmeasured

61% of enterprise AI projects were approved on projected ROI that was never actually measured after launch. Projects that define quantified success metrics upfront succeed 54% of the time; those that don't succeed only 12% of the time.

Projects Face Abandonment

Gartner projects that over 40% of agentic AI projects will be canceled by the end of 2027, and separately estimates 60% of AI projects lacking AI-ready data will be abandoned through 2026.

Governance Lags Automation

Only 21% of organizations deploying agentic AI report having a mature governance model in place — meaning most are automating decisions faster than they can explain or correct them.

This isn't an argument against AI. It's an argument against buying it blind. The vendors that can prove outcomes will earn the sale. The ones that can only prove activity are the ones that fail the fastest — often after the check has already cleared.

“It Completed the Task” Isn’t the Same as “It Got You Paid”

Most AI vendor pitches are built around task-completion metrics: claims processed, calls placed, codes assigned, denials “worked.” These numbers are easy to demo and easy to report. They are also, on their own, meaningless.

Task-completion metrics (what most vendors show you)	Outcome metrics (what actually matters)
Claims processed per hour	Claims that converted to payment
Calls or messages sent to payers	Denials actually overturned
Codes assigned or reviewed	Reduction in preventable denial rate
Tasks closed in the queue	Dollars collected per touch
“Automation rate”	Change in cost to collect

A system can process claims all day, work every denial in the queue, and close every task on a worklist — and if none of that activity is tied back to whether the claim was actually paid, the organization has simply automated its way to a faster version of the same invisible waste it already had.

That distinction — activity vs. outcome — is the single question every revenue cycle leader should hold onto through an entire AI sales cycle.

The Hidden Assumption Behind Most AI Pitches

Underneath almost every AI vendor pitch sits the same unstated assumption: removing a human from a task automatically creates value.

It doesn't. It only creates value if the task was worth doing in the first place.

- If 65% to 85% of the touches in a given revenue cycle workflow are already non-actionable — work that never changes a claim's outcome — then automating that workflow doesn't eliminate the waste. It just makes the waste faster and cheaper to produce.
- This is the real risk of task-completion AI at scale: cheaper mistakes, made faster, in higher volume — not better outcomes.
- Nowhere is this more visible than in claim intake. Front-end patient financial clearance — eligibility, authorization, registration — is responsible for over half of all denials industry-wide. An AI tool that speeds up a broken front-end process will speed up the denials it produces right along with it.

An AI vendor that can't tell you whether the work AI is automating was actionable in the first place isn't offering intelligence. It's offering acceleration — in whatever direction the process was already headed.

Why This Keeps Happening: The Data Gap Behind Most RCM AI

This isn't a failure of any single vendor. It's a structural gap that predates the current AI cycle by decades.

Confirms Work, Not Value

PM/EMR systems and the AI built on them can confirm a task was completed, not whether it mattered. There's no feedback loop connecting action to payment.

No Baseline, No Optimization

Without a real operational baseline, AI has nothing meaningful to optimize against. It just gets faster at the wrong thing.

The Infrastructure Gap

Only about 25% of executives are confident their data infrastructure can support scaling AI, and healthcare EHR integration keeps costing more than scoped.

Adoption Without Trust

Trust hasn't caught up to adoption. Half of healthcare leaders cite data quality and privacy as their biggest AI barrier, and 4 in 10 don't fully trust AI's results today.

> **Vendors aren't lying when they show you a task-completion dashboard. That's usually the only data their own system, or yours, has ever been built to produce.**

The Prerequisite for AI Success: Visibility Before Automation

"Measure first" isn't a nice-to-have step before an AI rollout. It's the non-negotiable prerequisite for one to work at all.

Classify Before You Automate

You cannot automate work you have never classified. If you don't know which touches are actionable and which aren't, an AI vendor can't either — they'll simply inherit your existing blind spot at machine speed.

The ROI You Can't Prove

You cannot prove ROI on a baseline you never established. Nearly two-thirds of enterprise AI projects are approved without a measured, quantified success metric — exactly why so many are quietly abandoned twelve months later with nothing to show a board except a bill.

Autonomy Isn't a Shortcut

The natural progression is sequential, not simultaneous: visibility first, automation second, autonomy third. Skipping to automation without visibility doesn't compress the timeline — it just moves the failure further downstream, where it's more expensive to find.

Organizations that measure operational effort before they automate it are the ones best positioned to tell — quickly, and in dollars — whether a given AI deployment is actually working.

The Vendor Vetting Framework

Four categories of questions to bring into every AI vendor conversation. If a vendor can't answer these directly, that's information too.

Outcome Measurement

- Does the platform tie specific AI-driven actions to payment outcomes — not just task completion?
- Can you see, claim by claim, which AI-touched claims were actually paid versus which just moved to a different queue?
- Is "success" defined as a task closed, or a dollar collected?

Baseline & Benchmarking

- Can the vendor show real before/after data from existing customers — not projected savings modeled from an industry average?
- Do they know your organization's current non-actionable touch rate, denial rate, and cost to collect before they propose a solution?
- Will they commit to measuring your baseline first, even if that delays the sale?

Transparency & Accountability

- What happens when the AI is wrong — is there an audit trail showing what data it used and what action it took?
- Who owns the error: the vendor, or the organization that deployed it? Get this in writing, not just in the pitch.
- Can a human reviewer override or correct an AI-driven action before it reaches a payer or a patient?

Scalability of Waste

- Could this tool make non-actionable work faster and cheaper instead of eliminating it?
- Is the AI targeting a process your own data shows is already high-value — or just the process that was easiest to automate first?
- What guardrails exist to prevent the AI from confidently repeating the same avoidable error at higher volume?

Red Flags vs. Green Flags

Built to be skimmable and shareable on its own - pin this to the wall before your next vendor call.

Before you move forward, ask:

1. What financial outcome will change?
2. How will we measure that change against our own baseline?
3. What happens—and who is accountable—when the AI is wrong?

Red Flags (Task-Completion AI)	Green Flags (Outcome-Driven AI)
Pitch leads with claims processed, calls made, or "automation rate"	Pitch leads with payment conversion, denial reduction, or cost-to-collect impact
ROI is projected from industry averages, not your own data	ROI is modeled from your organization's actual baseline
No visibility into which touches were actionable before automating them	Explicitly classifies actionable vs. non-actionable work before deploying
Vague or absent answer on who owns an AI-driven error	Clear, contractual answer on accountability and override
Success metric is "time saved" or "tasks completed"	Success metric is dollars collected or waste eliminated
Implementation timeline skips a baseline-measurement phase	Baseline measurement is a required first phase, not an afterthought
Vendor discourages requesting reference customers' real numbers	Vendor proactively shares before/after data from existing clients
Positions AI as a replacement for visibility	Positions AI as something that requires visibility to work

What Outcome-Driven AI Looks Like in Practice

SCENARIO A:

Task-Completion Deployment

A billing team deploys an AI tool to auto-draft appeal letters for every denial in the queue, regardless of denial reason. Appeal volume goes up. Staff hours per appeal go down. Six months later, the appeal overturn rate hasn't moved, because the tool was equally fast at appealing denials that were never going to be overturned as it was at appealing the ones that would.

SCENARIO B

Outcome-Driven Deployment

A different organization first classifies its denial-related touches by root cause and actionability. It learns a specific subset - denials tied to a handful of correctable, recurring errors - represents the large majority of preventable cost. It deploys AI narrowly against that subset, with the explicit goal of reducing the recurrence of those specific errors upstream, not just processing the resulting denials faster.

The difference isn't the sophistication of the model. It's what each organization asked the AI to optimize against. One organization measured first and gave its AI a real target. The other gave its AI a queue.

What Visibility-First Execution Looks Like

50%
less anticipated new A/R headcount

75%
fewer coordination-of-benefits denials

300%
increase in point-of-service collections

Next Steps

Evaluating an AI vendor doesn't have to mean becoming an AI expert. It means asking the same question about every claim, every dashboard, and every pitch: did this actually get us paid, or did it just look busy?

1

Start with your own data, not the vendor's demo

Before any AI conversation, know your organization's current non-actionable touch rate and denial rate. Without that baseline, every vendor ROI claim is unverifiable by definition.

2

Make the vendor vetting framework a required step, not a courtesy

Put the four categories above into your RFP or discovery call. A vendor's willingness to engage with these questions is itself a signal.

3

Sequence it deliberately

Visibility first. Automation second. Autonomy third. Organizations that skip step one are the ones funding next year's list of abandoned AI pilots.

MedEvolve's position on AI has been consistent: automation that doesn't know whether it got the organization paid just creates cheaper mistakes at scale. MedEvolve's Effective Intelligence® platform captures every human touch in the revenue cycle and ties it to financial outcome first — the exact baseline an AI vendor, including MedEvolve's own AI capabilities, needs in order to be evaluated honestly.

> ***"We are thrilled with the outcomes achieved with MedEvolve. Our leaders have tremendous confidence in the partnership."***
>
> — Heather Richards, CFO, Atlas Healthcare Partners

Sources

MIT NANDA Initiative, "The GenAI Divide" (2025)
MIT Sloan Management Review, enterprise AI ROI measurement study (2025)
Gartner, GenAI and agentic AI project abandonment forecasts (2024–2026)
BCG, "The Widening AI Value Gap" (September 2025); BCG AI Radar (2026)
Deloitte, agentic AI governance maturity research (2025–2026)
KLAS Research, 2025 healthcare AI adoption survey
McKinsey & Company, "Agentic AI: The Race to a Touchless Revenue Cycle" (January 2026)
Becker's Hospital Review / Forbes Technology Council, healthcare AI failure-rate commentary (2025–2026)
CAQH, 2025 CAQH Index
Experian Health, "3rd Annual State of Claims Survey" (2025)
MedEvolve customer results and testimonials, medevolve.com