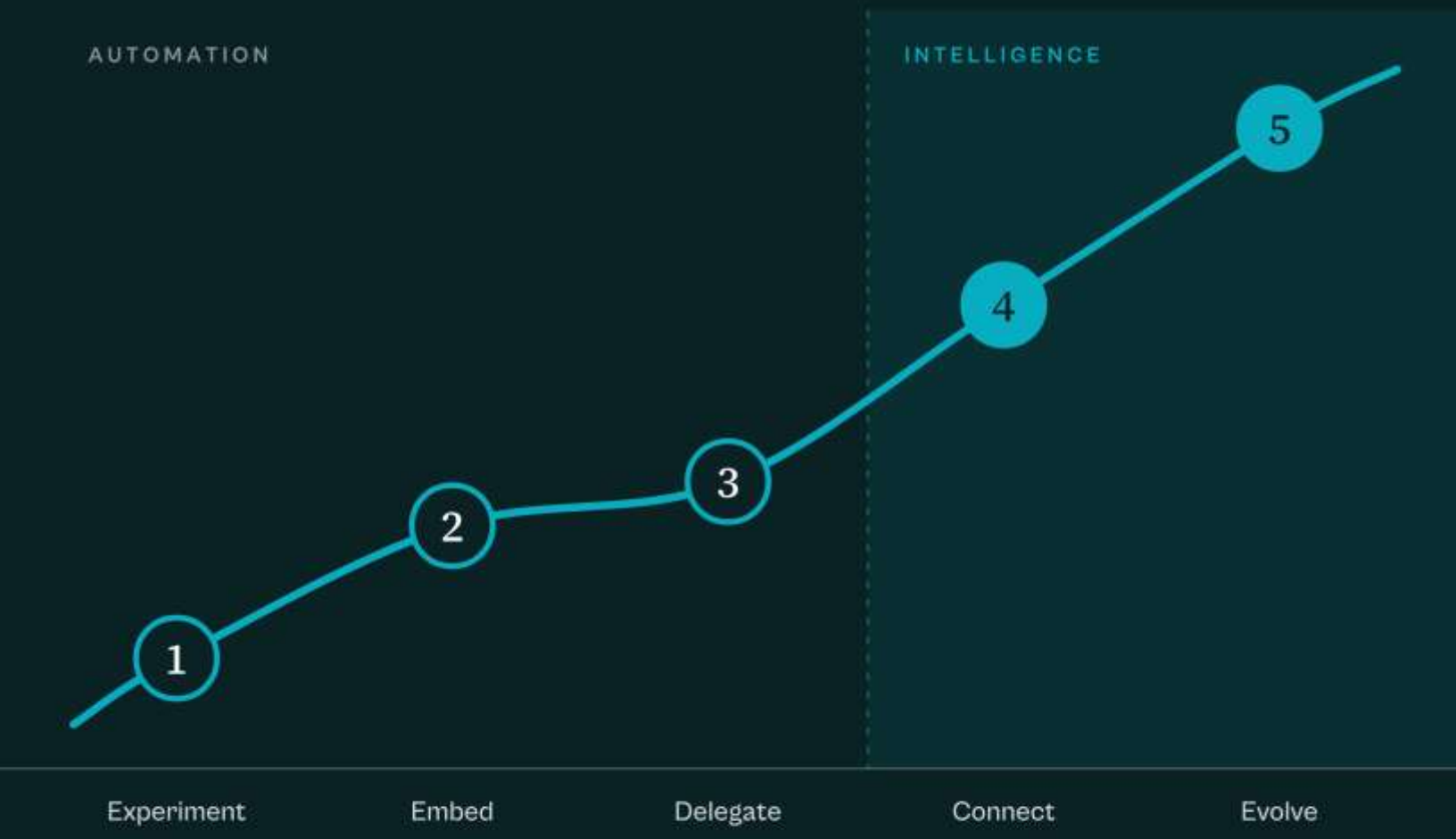
WHITE PAPER · PHARMA COMMERCIAL EDITION

AI that *Earns* its Place in Pharma Commercial

What to ask, where the value is and what to do before you pick your 2027 AI vendors

Rahul Agarwal SVP and GM, Healthcare and Life Sciences · November 2026



accellor

Contents

Executive summary 3

Introduction 4

1 Every AI firm now has agents. Ask what theirs have been trusted to do on their own. 5

2 The 6% problem: why pharma's AI pilots don't reach the P&L 9

3 Why content AI stalls at MLR review 12

4 What it takes for 300 scientists to trust an agent without a reviewer 15

5 Where to start 17

Appendix: What is in production and what is ready to build 18

About Accellor 18

Sources and notes 18

ABOUT THIS PAPER

This paper is for commercial excellence, commercial AI, analytics and digital leaders at pharma companies who are shaping 2027 plans. Every proof point comes from work in production, or is clearly labeled as designed and ready to build. Short on time? Read the executive summary, then Section 5, the five-move checklist.

Executive summary

Most pharma companies now use AI. Very few see it in their numbers. The difference is not the model: it is what the AI has been trusted to do, and the evidence behind that trust.

90%

of organizations now use AI regularly (McKinsey, 2026)

~6%

see significant value from it, 5%+ of EBIT (McKinsey, 2026)

70%

of AI value comes from changing how work gets done (BCG)

Key findings

1. **Ask what an AI system has been trusted to do on its own, and what evidence earned that trust.** Demos prove nothing about autonomy (Section 1).
2. **Most pharma companies use AI; about 6% see it in their numbers.** Programs stall between Stage 2 and Stage 3, and roughly 70% of the value comes from redesigning how the work gets done, not from the model (Section 2).
3. **In medical, legal and regulatory review, put the agent in front of the reviewer, not in place of one.** Pre-review speeds the queue; the control point stays human (Section 3).
4. **Consequence sets the ceiling on autonomy.** One governance policy for every agent will fail; evidence on file is what lets an agent act alone (Section 4).
5. **Start with an inventory, one well-chosen workflow and the evidence to scale it** (Section 5).

Introduction

I spent two days at the Reuters Events Pharma Customer Engagement conference in Philadelphia in October. Almost every booth led with the same word: agents. Almost every commercial leader I spoke with had the same question: which of these will actually show up in our numbers?

This paper is my answer. It starts with the one question I'd ask any AI partner before signing for 2027. It then looks at why so few pharma AI programs reach the P&L, where medical, legal and regulatory review quietly caps what content AI can deliver, and what it took for more than 300 scientists at a global pharma leader to trust an agent without a reviewer. It ends with the moves I'd make now. For most teams, 2027 spend is already committed, but the vendors and solutions often aren't picked yet.

Every proof point comes from work in production, or is clearly labeled as designed and ready to build. Use it to pressure-test your 2027 shortlist.

THE PRINCIPLE BEHIND THIS PAPER

Autonomy is earned, and capped by consequence. A system gets more freedom as it proves reliable, but never more than the cost of a mistake allows. Internal, low-risk work can run inside set limits. HCP- and customer-facing work needs a named person to approve. Patient safety, product quality and regulatory submissions always keep the decision with a named person.

Rahul Agarwal
SVP and GM, Healthcare and Life Sciences, Accellor
November 2026

SECTION 1

1 Every AI firm now has agents. Ask what theirs have been trusted to do on their own.

The one question I'd ask any AI partner before signing for 2027, and three follow-ups that separate real agents from relabeled software.

Walk the floor of any pharma conference this year and you'll hear the same word in almost every booth: agents. Product companies, analytics firms and AI start-ups all lead with some version of "agentic AI in production." To be fair, so do we.

And that's the problem: everyone uses the same word to describe very different solutions, from relabeled chatbots to systems that truly act on their own.

Gartner has a name for this: agent-washing, the rebranding of chatbots, robotic process automation and assistants as agents. In June 2025 it estimated that only about 130 of the thousands of vendors claiming agentic AI are the real thing. It also predicted that more than 40% of agentic AI projects will be canceled by the end of 2027, because of rising costs, unclear value or weak risk controls.[1]

I trained as a physician, and I've spent the last decade working on AI applications in healthcare. In our industry, a mislabeled system can touch an adverse event, a promotional claim or a regulatory submission. Buying the wrong thing costs more here than almost anywhere else.

1.1 The question I'd ask first

Demos are easy. Almost any model can draft a pre-call plan or summarize a publication on stage. So, I've stopped asking what a vendor's AI can do. I ask a harder question: **what has your AI been trusted to do on its own, in production, for a real client, and what evidence earned that trust?**

We hold ourselves to the same question. At Accellor, trust is earned in five levels, and each has a different bar.

FIGURE 1 Five levels of trust: autonomy is earned on evidence

Each level has a higher bar. A system can be pulled back a level the moment its evidence slips.



Consequence sets the ceiling: the highest level a system may reach depends on what happens if it is wrong.

Two of our own agents show what earning a level looks like. At a global pharma leader, a governed knowledge agent answers questions for more than 300 R&D users on its own, read-only (Section 4 tells that story). At a global medical-device maker, a purchase-order-to-invoice agent now acts alone within limits. It earned that by holding 100% accuracy under human review, and exceptions still go to a person. Either one can be pulled back a level the moment its evidence slips.

1.2 A good solution is more than agents

The second thing I check is what kind of AI I'm actually being sold. There are four kinds, and the difference is who controls what happens next.

TABLE 1 **Four kinds of AI, and who decides what happens next**

Kind	Who decides	How to recognize it	Example
Infrastructure	Rules decide	Fixed rules and code. The same input gives the same output; no model decides anything.	Routing, validation, integrations
Assistant	A person decides	The model writes the response. A person or the calling system decides the next step.	Drafting, summarizing, Q&A
Bounded task agent	The model owns one goal	It owns one narrow goal and chooses its own steps and tools to reach it.	Pre-call plan, claims pre-check
Truly agentic	The model plans the work	It plans, picks tools or other agents, recovers from failure and decides when it's done.	Multi-step work with no fixed path

The further down the table, the more control the model has, and the more evidence you need before trusting it.

In the abstract, none of these is better than the other. The best solutions I've seen use rules where the answer is known, predictive models where something has to be scored or forecast, and language models to draft and explain. They keep the planning agent for multi-step work where the path can't be fixed in advance. So ask every vendor which of the four they're actually selling. If the answer is "agents" across the board, you won't know what you're paying for or who is accountable when it's wrong.

1.3 Three follow-up questions

1. **Who decides the next step:** your configured rules, a person or the model?
2. **What happens when it fails partway through a task?** Does it recover, stop or escalate, and who sees the failure?
3. **What can it do without a person approving it,** and how is that limit enforced and logged?

Then I ask for the evidence: the test set and its pass marks, the audit trail for a sample of real outputs, and the rule that demotes the system when performance drops. If a partner can't answer precisely, they're probably selling an assistant or infrastructure under an agent label. That may still be the right product. It should just be priced and governed as what it is.

1.4 Consequence sets the ceiling

One more test is specific to life sciences. Autonomy is earned, but it's also capped by consequence. Internal, low-risk work like search and meeting notes can run inside set limits. HCP- and customer-facing work needs a named person to approve before anything goes out. Patient safety, product quality and regulatory submissions always keep the decision with a named person.

Commercial teams often assume their AI is all low-risk, but it isn't. A field conversation can surface an adverse event, or a medical question a rep must not answer. The right system catches those moments and routes them to pharmacovigilance and medical affairs without acting on them.

> *"Stop asking what the AI can do in a demo. Ask what it has been trusted to do on its own, and what evidence earned that trust."*

NEXT *Knowing what you're buying is the first step. The next question is why so few pharma AI programs ever show up in the numbers.*

SECTION 2

2 The 6% problem: why pharma's AI pilots don't reach the P&L

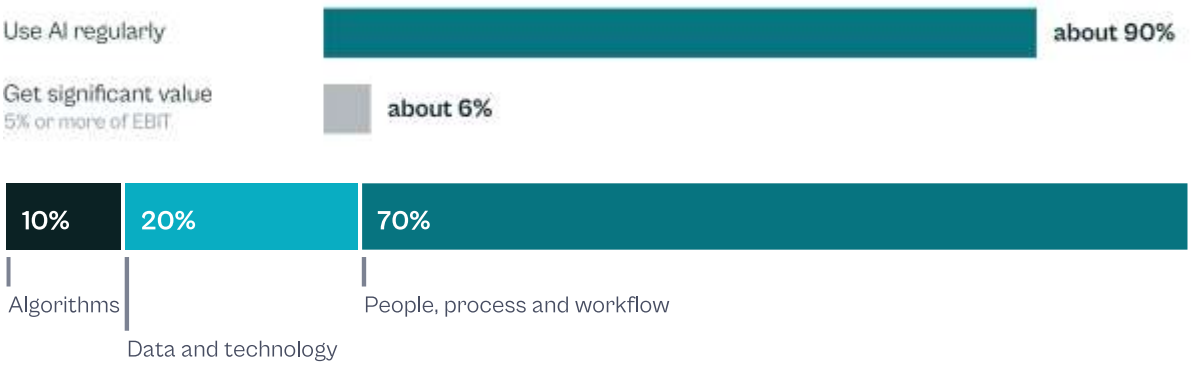
Almost every pharma company uses AI, but very few see it in their numbers. What I heard in Philadelphia, and what the companies that get value do differently.

One statistic frames the challenge. McKinsey's 2026 State of AI research found that roughly 90% of organizations now use AI regularly, but only about 6% get significant value from it, defined as 5% or more of EBIT.[2]

Pharma is no exception. Most commercial teams I talk to have more AI pilots than they can count. Very few can point to a line on the P&L that moved.

FIGURE 2 **Most teams use AI. Few see it in their numbers.**

Share of organizations, and where AI value comes from



Sources: McKinsey, The State of AI 2026;[2] BCG 10-20-70 rule of thumb.[3]

In Philadelphia, the sessions rarely dwelled on models or platforms. Three themes came up consistently:[5]

Silos	AI deployed in silos amplifies the silos. Each function gets its own tool, and nothing connects.
Data	AI-ready data is the bottleneck. The model is rarely what holds a program back.
People	Adoption depends on people, process and incentives far more than on the technology.

Those three themes explain the gap between the 90% and the 6% better than any vendor slide I've seen.

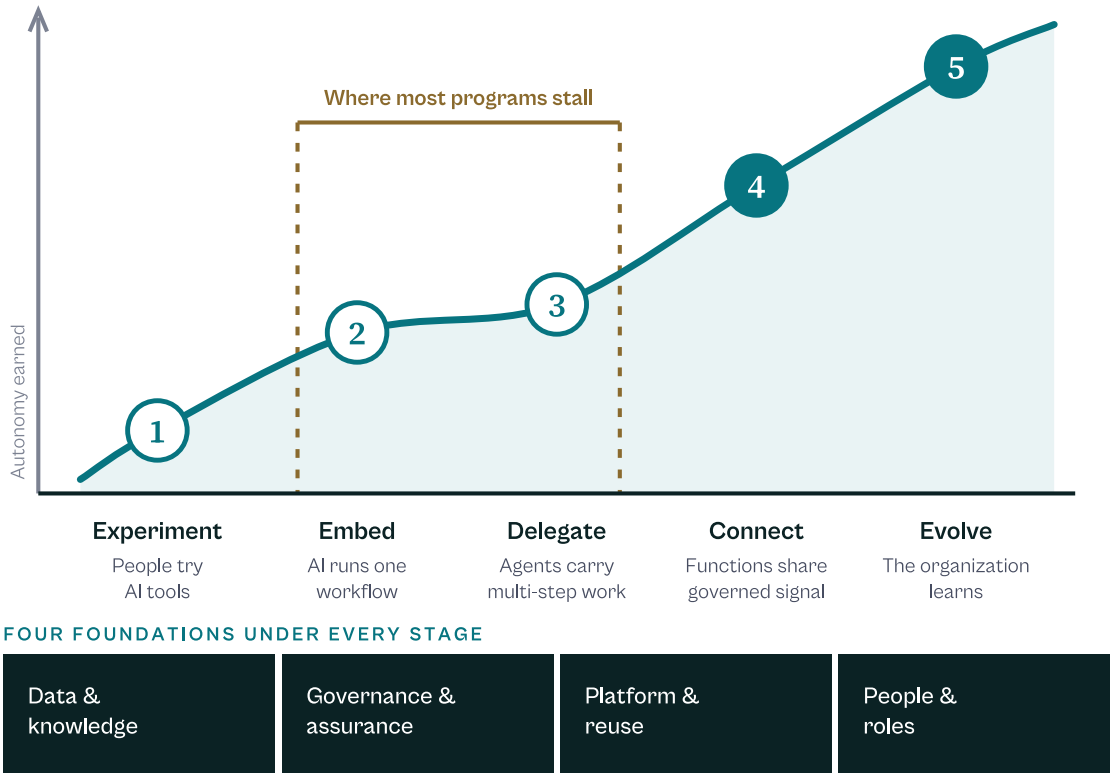
2.1 The plateau between Stages 2 and 3

We describe AI adoption in five stages: Experiment, Embed, Delegate, Connect and Evolve. In the first three, AI does more of the work inside a function. In the last two, functions start sharing what AI learns, and the whole enterprise learns.

Most pharma companies are stuck between Stages 2 and 3. Each team has its own AI, but a finding in one function still reaches the next by email and meetings. That's progress, but it isn't transformation, and it's why the value captured stays small.

FIGURE 3 Most pharma companies plateau between Stage 2 and Stage 3

Place each function on the curve separately: a company can be at Stage 3 in commercial operations and Stage 1 in medical.



2.2 Seventy percent of the value lies in how the work gets done

BCG's 10-20-70 rule of thumb puts it plainly. About 10% of AI value comes from algorithms, 20% from data and technology, and 70% from changing how the work gets done.[3] McKinsey's data points the same way: AI high performers are roughly three times as likely as everyone else to have fundamentally redesigned their workflows.[2]

I find that encouraging for commercial leaders. **Workflow and operating model are things you control.**

2.3 What redesign looks like in commercial

Take field engagement. I'd start with call preparation, because it's high-volume and needs real judgment about the HCP.

Stage 2 · Embed	AI drafts a pre-call plan from approved content and CRM data, and the rep reviews it.
Stage 3 · Delegate	An agent prepares the week's plan, drafts call summaries and flags possible adverse events heard in the field, sending a draft report to safety for the safety team to decide.
Stage 4 · Connect	Medical and supply signals reach territory planning the same day instead of the next quarter.

TWO THINGS NEVER CHANGE

The conversation with the physician stays with the rep at every stage. And AI never creates claims. It works only from reviewed content, so the review process stays the control point.

> *"The model is rarely what holds a program back. Disconnected data, unclear owners and unchanged workflows are."*

2.4 Data is the bottleneck nobody budgets for

In Deloitte's 2026 survey, only 34% of life sciences and health care companies rated their data management as highly prepared for AI, a lower share than the year before.[4] An agent acts on what it can see. If customer, product and account records disagree across systems, it acts quickly on a flawed picture of the business.

The fix isn't a data program that runs for two years before anything ships. It's fixing the data your first workflow depends on, as part of building that workflow: one authoritative record for each HCP and account, access through governed interfaces, and a named owner who corrects errors.

NEXT *Content review is where these gaps show up first, and where the pressure is greatest right now.*

SECTION 3

3 Why content AI stalls at MLR review

Content teams have AI, review teams have queues. How agents can speed medical, legal and regulatory review without touching the control point.

Generative AI has made content faster to produce, but it hasn't made content faster to approve.

Most commercial teams now generate more drafts, more variants and more channel-specific versions than ever. Every one of them still has to pass medical, legal and regulatory review with the same reviewers, same checks, working longer queues.

McKinsey estimates that generative AI applied to MLR could make content approval two to three times faster and cut review time by 50% to 70%.[2] Few companies are seeing that yet. In my view, that's because most of the investment has gone into generating content, not reviewing it.

3.1 Why generation alone makes it worse

Volume	Volume rises faster than review capacity, so time to market gets longer, not shorter.
Drift	AI-written drafts raise the risk that a claim drifts from the approved language, so reviewers check harder.
Rework	Assets come back for the same corrections round after round, because nothing learns from the last review.

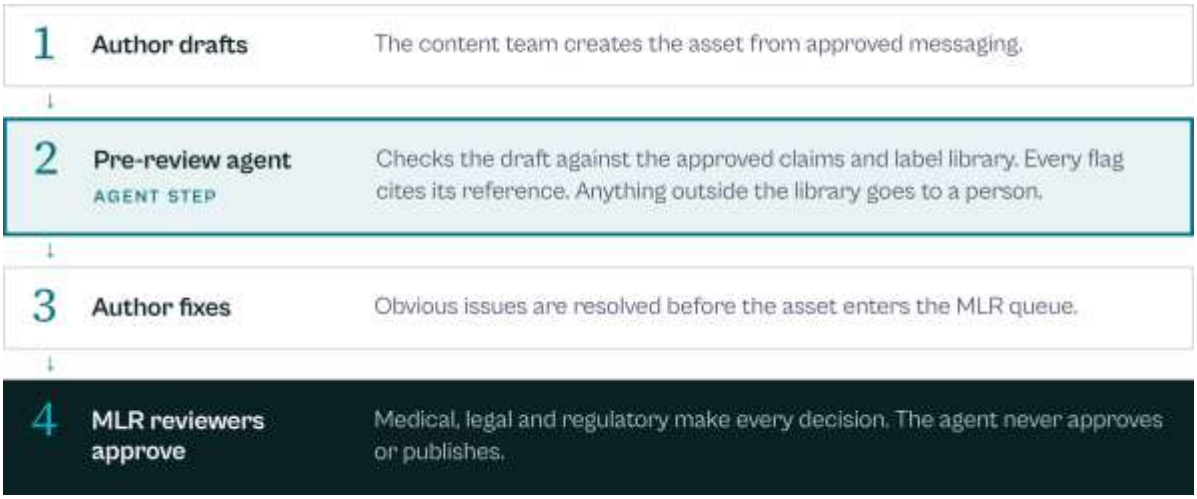
3.2 Pre-review, not auto-approval

The pattern I believe in puts an agent in front of the reviewer, not in place of one. Before a draft reaches the MLR queue, the agent checks it against the approved claims and label library. Each flag cites the reference it relied on, so the reviewer sees exactly why it was raised. Anything outside the library always goes to a person. **Reviewers approve. The agent never does.**

Done this way, reviewers spend their time on judgment calls instead of mechanical checks. Authors fix the obvious issues before anything hits the queue.

FIGURE 4 **The pre-review agent sits in front of the reviewer**

Reviewers spend their time on judgment calls; authors fix the obvious issues before anything reaches the queue.



Status: designed and scoped, ready to build.

> *"Put the agent in front of the reviewer, not in place of one."*

3.3 Where this sits on the trust ladder

Promotional content is HCP- and customer-facing work, so a named person approves before anything goes out. Some elements, like fair balance and safety information, carry higher stakes still. The ceiling is clear: the agent may suggest and draft; it doesn't approve or publish.

That ceiling doesn't rise as the system matures. What improves is the quality of the flags, and the time reviewers get back.

3.4 What has to be true before you start

Library	A current, structured claims and label library, with versions and status for every reference.
Traceability	Reference documents that are labeled and retrievable, so every flag can be traced to a source.
Learning	Review outcomes captured in a usable form, so the system can learn from reviewer corrections over time.
Audit	An audit trail that records the reviewer, the version and the evidence for every decision.

3.5 Where it fails

I see three failure modes:

1. **An out-of-date claims library,** which makes the agent confidently wrong.
2. **Flags that reviewers can't trace to a source,** which destroys trust within a week.
3. **Measuring speed before quality,** which rewards rubber-stamping.

Track review cycles per asset, days to approval, first-pass rate and reviewer override rate together.

WHERE WE ARE

We've designed and scoped MLR pre-review as a pattern, ready to build. We'd build it the same way as the agents we already run in production for a global pharma leader: a test set with pass marks before every change, locked versions, a full audit trail and a named owner. It's a pattern, not a product, designed around your review process rather than replacing it.

NEXT *Every pattern so far depends on one thing: evidence that the system deserves the autonomy it has. The next section shows what that evidence looks like in practice.*

SECTION 4

4 What it takes for 300 scientists to trust an agent without a reviewer

A governed knowledge agent at a global pharma leader answers R&D questions on its own. The model was the easy part. Here's what earned the trust.

At a global pharma leader we work with, more than 300 R&D users ask a governed agent questions in plain language, across portfolio, clinical and finance data. It answers on its own. There's no reviewer in the loop.

In an industry that's cautious about AI, for good reasons, that's unusual. And it didn't happen because of the model.

300+
R&D users asking questions in plain language, no reviewer in the loop

Up to 50%
faster portfolio decision cycles

Up to 90%
less analyst effort on portfolio reporting

4.1 The problem we started with

The company had more than 150 clinical assets across siloed therapeutic areas. Research, clinical and portfolio data couldn't be seen together, so answering a portfolio question meant analysts assembling reports by hand. Decisions waited on reporting cycles.

Today, governed agents answer portfolio questions directly, under one access policy, so each person sees only what they're permitted to see. Portfolio decision cycles are up to 50% faster, and analyst effort on portfolio reporting is down by as much as 90%.

4.2 What earned the trust

The agent acts alone because it does read-only internal work, and because the evidence for every one of these is on file:

TABLE 2 **The evidence on file**

Intended use and risk	What it may do, and the harm if it's wrong. That sets its tier.
Test protocol	An 800-question regression set with pass marks runs before every change.
Locked versions	Model, prompts and tools are versioned; changes go through change control.
Cited answers and audit trail	Every answer traces to its source, with version and evidence recorded.
A named owner	One accountable person, not a committee.
Automatic demotion	Below the pass mark, it drops a tier until it earns its way back.

In medicine, a decision is trusted because there's evidence behind it and a named person accountable for it. I don't see why AI should be held to a lower standard.

> *"The agent acts alone because the evidence is on file, not because the model is impressive."*

4.3 One governance policy for every agent will fail

Most companies govern AI one of two ways. Tight, uniform controls block safe, low-risk uses and push people toward unapproved tools. Loose, uniform controls expose quality and patient-safety work to risks nobody has sized. Gartner argued in May 2026 that applying uniform governance across AI agents will lead to enterprise failure. In its 2025 research, only 13% of IT leaders strongly agreed they had the governance to manage agents.[1]

The alternative is the rule from Section 1: match controls to consequence. The same company can, and should, run all three at once.

TABLE 3 **Match controls to consequence**

Consequence	Examples	Default ceiling
Internal, low-stakes	Research Q&A, search, meeting notes	Runs alone within limits once it passes its evaluation
HCP- and customer-facing	Pre-call plans, promotional content	Drafted by AI; a person approves
Safety, quality, regulatory	Adverse events, submissions	Acts only with approval in a validated workflow; the decision stays with a named person

SECTION 5

5 Where to start

Five moves to make before you pick your 2027 vendors.

None of this requires a multi-year program before anything ships. It requires knowing what you have, choosing well, and funding the evidence from the start.

1 **Inventory.** Build one inventory of every AI tool and agent you use or are buying, with a named owner and its permissions.

2 **Classify.** Classify each one by kind (infrastructure, assistant, task agent, truly agentic) and by the trust level it holds today. Then match its governance to the consequence, not to the vendor's label.

3 **Place on the curve.** Place each function on the five stages separately. A company can be at Stage 3 in commercial operations and Stage 1 in medical.

4 **Pick one workflow.** Pick one Stage 2 workflow that carries real expert judgment. Give it a named owner, fix the data it depends on, and measure it against today's baseline.

5 **Fund the evidence.** Require a test set with pass marks before anything is built and set the demotion rule in advance. Decide in 90 days whether to raise that workflow's autonomy or hold it.

> *"Governance done this way isn't a brake. It's what lets the tenth agent ship in a fraction of the time the first one took."*

BRING ME YOUR 2027 AI SHORTLIST

Send me the two AI workflows or vendor proposals at the top of your 2027 list. In a 30-minute working session, my team and I will tell you what kind of AI each one really is, what level of trust it has earned, and what evidence to ask for before it goes further. You'll leave with a one-page view you can take to your steering committee.

Or go deeper in 45 minutes: place your commercial functions on the five stages, walk us through one asset's MLR journey, or apply our consequence-tier rule to five of your active AI workflows.

Book a time with Rahul

Appendix: What is in production and what is ready to build

Capability	Status	Where
Governed knowledge agent for 300+ R&D users, read-only, no reviewer	**In production**	A global pharma leader
Portfolio decisions up to 50% faster; analyst reporting effort down up to 90%	**In production**	A global pharma leader
Purchase-order-to-invoice agent acting alone within limits (100% accuracy under review)	**In production**	A global medical-device maker
MLR pre-review agent	Designed and scoped (ready to build)	Accellor pattern

About the author

Rahul Agarwal is SVP and GM, Healthcare and Life Sciences, at Accellor. He trained as a physician and has spent the last decade working on AI applications in healthcare.

About Accellor

Accellor has worked with healthcare and life sciences organizations for more than eight years and on more than 70 projects. It is a Salesforce Summit Partner and a Microsoft Solutions Partner for Data & AI, with teams in the United States, the United Kingdom and India.

Sources and notes

1. Gartner: agent-washing estimate and agentic AI project cancellations (June 2025); uniform AI agent governance (May 2026); IT leaders' agent governance readiness (2025).

2. McKinsey & Company: The State of AI 2026; generative AI in medical, legal and regulatory review.

3. Boston Consulting Group: 10-20-70 rule for AI value.

4. Deloitte: 2026 life sciences and health care survey on data readiness for AI.

5. Reuters Events Pharma Customer Engagement USA, Philadelphia, October 2026: session themes.

Client examples are described without names. Capabilities marked "designed and scoped" are patterns ready to build, not products in production.

© 2026 Accellor. All rights reserved. · accellor.com