RAiL

Responsible AI
in Life Sciences

REPORT 2026:

Defining an Agentic AI Trust Benchmark

ACTO



Table of Contents

Executive Summary 02

Section 01
Research Methodology 03

Section 02
The State of AI In Life Sciences 09

Section 03
A Shared Vocabulary 13

Section 04
What We Heard: Six Emerging Themes from the Field 16

Section 05
From Themes to a Standard: The RAIL Score 36

Section 06
The Agentic AI Playbook: Where to Start, What to Build 42

Section 07
Redesigning Work: The Human-Centric Mandate 46

Section 08
The First 90 Days 49

Section 09
Grading A Vendor Before you Buy 51



EXECUTIVE **SUMMARY**

Agentic AI is moving into life sciences commercialization faster than the industry has developed a consistent way to evaluate it. To understand what leaders need from these tools and systems, we spent the first half of 2026 listening to the people responsible for putting them to work.

Through the **Responsible AI in Life Sciences (RAIL)** roundtable series, we spoke with over 80 leaders across North America and Europe in commercial, medical affairs, patient services, IT, and regulatory roles at pharmaceutical, biotech, and medical device companies to gather candid perspectives on responsible AI in commercialization. This insight was expanded with an online survey distributed to the same audiences profile, garnering over 100 qualified respondents.

We found that the debate has already moved past whether to use agentic AI or not. The potential impact of agentic AI to improve both upstream and downstream commercial work is understood. Instead, leaders were wrestling with a harder, more practical question: **which agentic AI solutions are trustworthy enough to scale?** They could see the potential for business and patient impact. But they also needed to know how to validate vendor claims, maintain oversight, and manage risk as AI agents take on field and patient-facing work.

Today, there is no shared language for distinguishing a robust agentic AI system from a white-labelled chatbot, or an agreed bar for what "responsible AI" means in practice. That makes it difficult for Life Sciences teams to compare vendors' claims and assess the risks.

This report offers a way to make those judgments. The practical question is not, "Is AI perfect?" but, "Is it safer, better, faster, or more effective than how the work is done today?" Answering it requires clear governance, validation criteria, risk thresholds, and business cases suited to a highly-regulated industry.

The concerns and requirements leaders raised shaped the **RAIL Score**: a rubric covering the six components of responsible agentic AI against defined criteria. It produces a single, comparable way to assess any agentic AI system or tool — moving the conversation from vendor claims and broad principles to evidence and measurable standards.

Our hope is that leaders use the RAIL Score with their teams to assess any AI solution they use or are considering—so they can see where it stands, what needs to improve, and whether it is ready to scale in Life Sciences.

-Parth Khanna | CEO, ACTO



SECTION 01

RESEARCH **METHODOLOGY**

How the study was conducted

This research combined three methods to gather both depth and breadth of perspective on responsible AI in Life Sciences, focused on commercialization. Perspective was gathered from leaders at pharmaceutical, biotech, and medical device companies from commercial, medical affairs, patient services, and market access, as well as the IT and governance functions behind them.

At a glance:

08	**08**	**80**	**02**	**135**
roundtables	cities	leaders	regions	survey respondents

1

Executive **Roundtables**

80 senior leaders

hosted at in-person roundtables across **8 cities** in North America and Europe with discussions focused on human-centric change, safety, governance & compliance, and AI use cases & impact.

2

Online **Survey**

135 respondents

across the industry from North America and Europe answered **14 multiple-choice questions** about Responsible AI in Life Sciences, fielded June–July 2026.

3

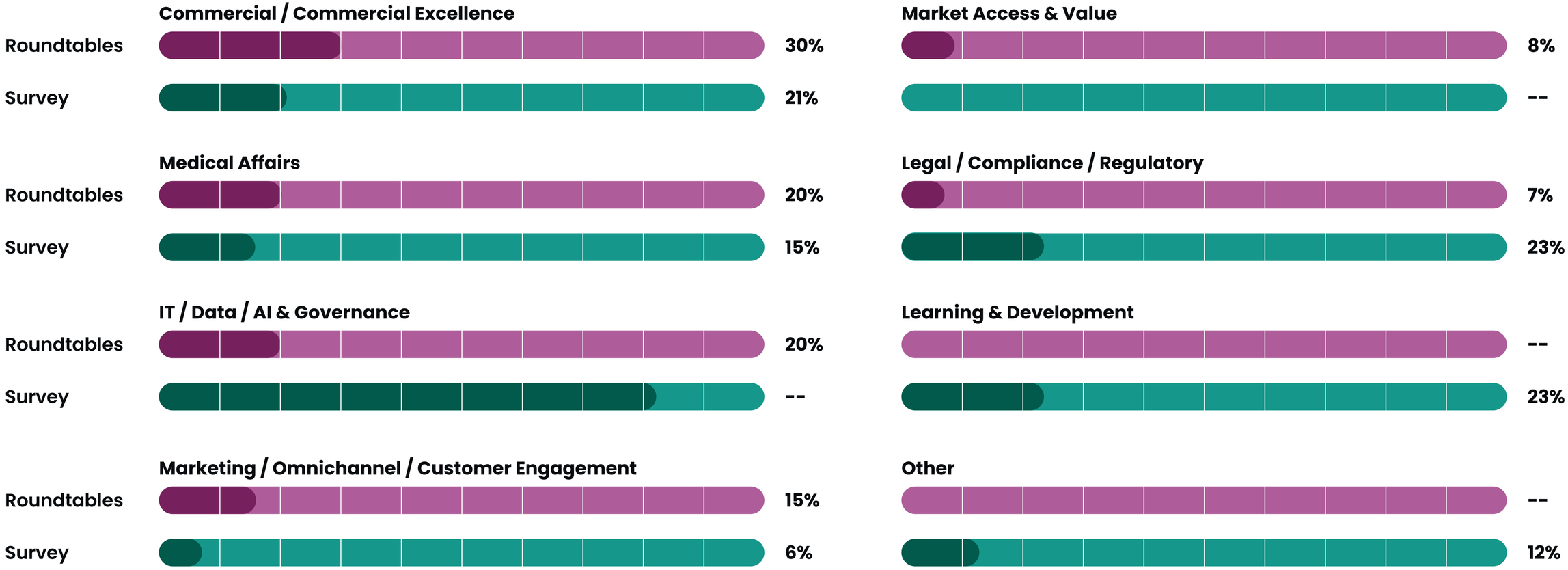
Source **Synthesis**

Coding of all **8 roundtable transcripts** into recurring themes were cross-referenced against the **online survey results** and existing standards (ISO/IEC 42001, EU AI Act).

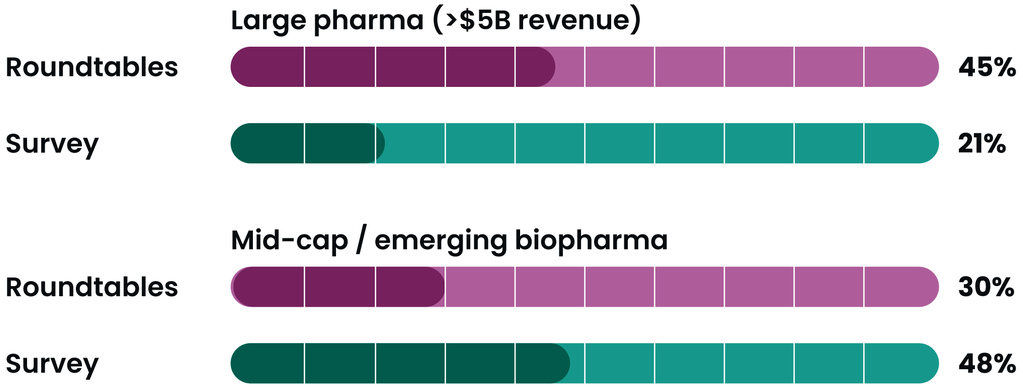
WHO **TOOK PART**

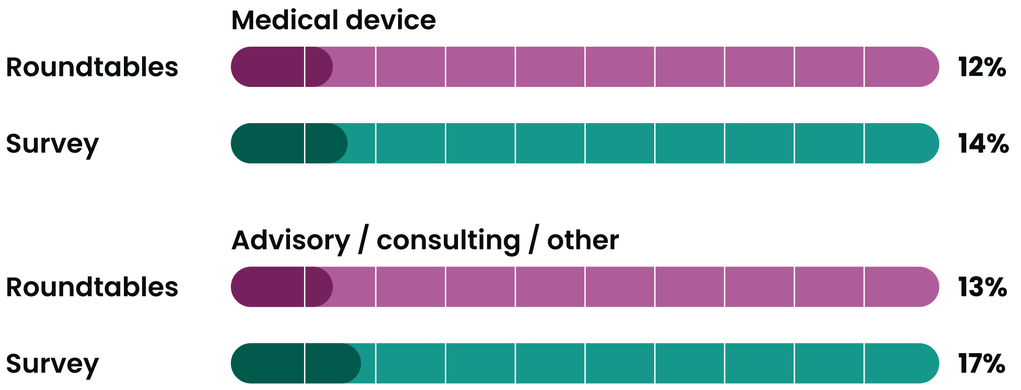
The roundtables and the survey convened senior leaders spanning large pharma, mid-cap and emerging biotech, and medical device firms.

By Function

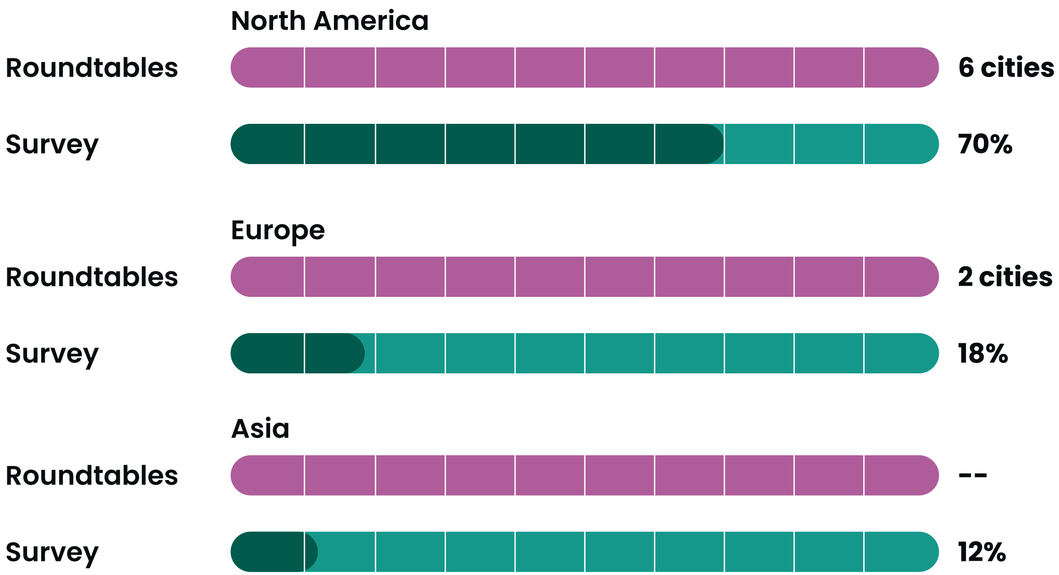


By Industry

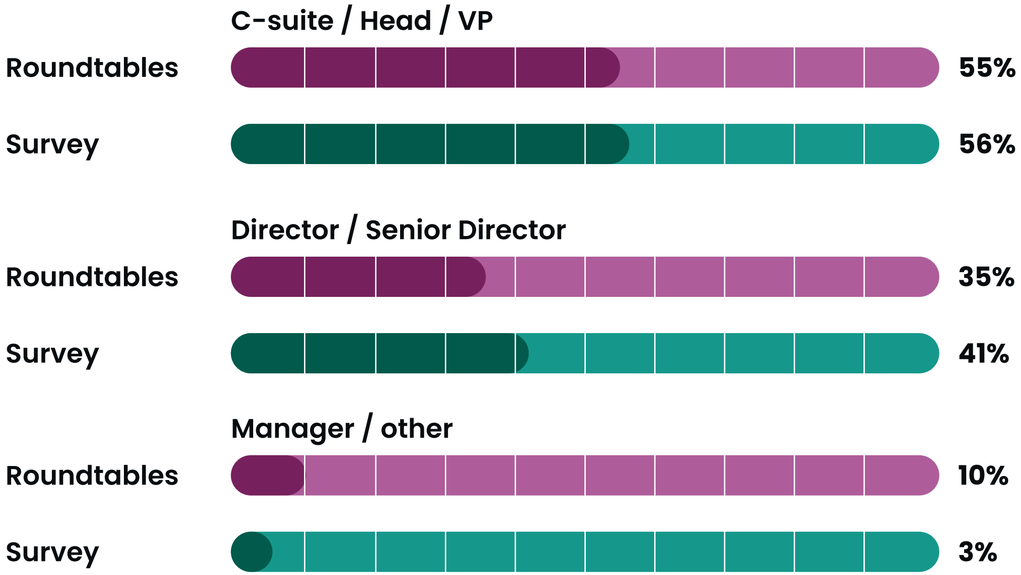




By Region



By Role



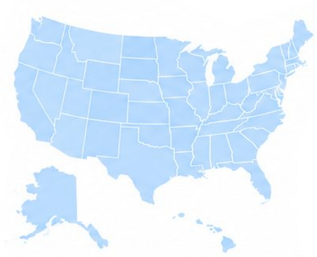
WHERE WE **CONVENED**

United Kingdom



GILEAD
AstraZeneca
MSD
SERVIER
IPSEN
Incyte
VIATRIS
BESINS HEALTHCARE
abbvie
Roche Diagnostics
astellas
BeiGene

United States of America



Genentech A Member of the Roche Group | Roche | Takeda | GILEAD
VERTEX
Johnson&Johnson MedTech
AstraZeneca
NOVARTIS
novo nordisk
Boehringer Ingelheim
abbvie
astellas
AMGEN
Biogen
sanofi
SOLENO THERAPEUTICS
Emalex
Johnson&Johnson
BIOMARIN
Sangamo THERAPEUTICS
CACTUS LIFE SCIENCES
Incyte
Boston Scientific
Deloitte.
Bristol Myers Squibb
Pfizer
HELSINN
Lundbeck
Johnson&Johnson
CSL
Telix
Revolution Medicines
arcutis
PACIRA BIOSCIENCES, INC.

Canada



sanofi
VIATRIS
Takeda
GILEAD
AstraZeneca
Boehringer Ingelheim
Thermo Fisher SCIENTIFIC
bdc

Switzerland



NOVARTIS
Roche
teva | mepha
ultragenyx
novo nordisk
BAYER
SUMITOMO
evinova



SECTION 02

THE STATE OF AI **IN LIFE SCIENCES**

The conversation about AI in life sciences is being pulled in two directions at once.

On one side is a hype cycle near its peak. Every vendor's product is "agentic," every deck promises autonomy, and the distance between the demo and the deployed reality has never been greater.

On the other is a real and rising caution: about job security, about systems that act without being watched, about putting anything near a patient or a regulator that can't explain itself. Leaders are being told to move fast and to be careful in the same breath, and the market has handed them no instrument that does both.



Chicago Roundtable

"Every industry right now is pausing and saying: what does this mean to us — this AI?"

From The Survey

Are you concerned about the use of AI at your organization?



How would you rate your level of comfort with AI? (1 = low, 10 = high)



N = 135

Personal comfort runs high. Yet more than half remain concerned about how AI is being used inside their own organization.

Leaders consistently shared three pressure points driving the urgency behind AI adoption:

- There is **more to commercialize**: AI is already transforming drug discovery, and now quickly moving into unlocking efficiency gains in drug development and clinical trials.
- There are **tighter budgets**: “Do more with less” is now a mandate.
- There is **strong pressure to adopt AI without a clear strategy**: Executive enthusiasm is outpacing clear plans for managing risk and proving ROI.

In this environment, agentic AI stops being a curiosity and becomes a practical lever for capacity, quality, and speed. This is why getting the discussions around future of work and the role that AI agents will play in the workplace have become a priority.



Princeton Roundtable

"Every single time we’ve had a breakthrough innovation, it gives human beings more agency and it gave us more ability.”

From The Survey

Are you currently involved in an AI project at your organization?



Do you believe there are clear industry standards for responsible AI in Life Sciences?



N = 135

Two in three leaders are already working on AI projects. More than half say there is no clear standard to build against.

What makes Life Sciences different is that the adoption question cannot be separated from trust, patient safety, compliance, and human judgment. A marketing-copy generator that hallucinates is an embarrassment; an agent that puts a wrong dose, an off-label claim, or an unmonitored message into a customer-facing workflow is a regulatory and patient-safety risk. **This is the healthcare equivalent of confidently wrong information reaching someone who acts on it.** The bar is simply higher here than in generic enterprise software.

Existing standards provide useful scaffolding, but they do not fully address the gap. ISO/IEC 42001 offers a certifiable management-system framework for governing AI across its lifecycle, and the EU AI Act classifies AI by risk tier and imposes real obligations on high-risk uses. But neither was written for the specific reality of Life Sciences customer-facing teams. For instance: a field sales rep answering a dosing question from an HCP, an MSL summarizing a new insight from a scientific exchange with an HCP, or a case manager supporting a patient program.

A bounded, industry-specific bar for what responsible agentic AI should look like in Life Science is missing and that absence is what all conversations kept circling.

This report is structured around the three questions that anchored roundtable discussions, and continue to shape the market:

01

Where can agentic AI create value in Commercialization?

02

How must safety, governance, and compliance be handled?

03

How is work redesigned so humans remain central?



SECTION 03

A SHARED **VOCABULARY**

“Agentic” is used loosely, often to describe very different kinds of AI. The clearest way to distinguish them is by autonomy: how much of the work the AI can carry out without human direction, and where control shifts from the user to the system. This distinction matters because labels can mislead.

A vendor may market a copilot as an “autonomous agent,” while an in-house tool may already be acting as an agent without being recognized —or governed—as one.

Types of AI Solutions

Prompt Tool	Copilot	Agent	Agentic AI system

TYPES OF **AI SOLUTIONS**

	Prompt Tool	**Copilot**	**Agent**	**Agentic AI system**
What is it?	You ask, it **answers**. A better search box.	**Drafts and suggests** inside a task you're already doing.	Completes a single task or workflow end to end. A **one-trick pony**: powerful, narrow, one job.	An orchestrator directing multiple sub-agents toward a complex goal while adhering to guardrails. A super-agent "**queen bee**", coordinating worker-agents.
Who's in control?	The user, at every step.	The user, with suggestions.	The agent, within set bounds.	The system, under governance.
Typical Life Sciences use cases	Summarize a clinical paper; draft a first-pass email.	Suggest edits to a call summary; recommend next-best content.	Compile pre-call prep for an HCP visit; log and tag an interaction.	Coordinate a product launch: prep, content, compliance checks, follow-ups.



AI INVENTORY WORKSHEET

Put the vocabulary to work

Download now

Before deciding what AI to build or buy, map what is already in use. This worksheet helps you classify each AI tool by autonomy and spot gaps in ownership or data boundaries.



SECTION 04

WHAT WE HEARD:
SIX EMERGING THEMES FROM THE FIELD

Across the roundtables, six themes emerged. The first three themes describe how responsible AI should be **designed and deployed:** around patients and human judgment, grounded in the right context, and introduced through bounded use cases with measurable value.

The remaining three themes describe what organizations need to scale it: governance that continues beyond initial approval, value measured against the current human baseline, and thoughtful workforce redesign.

Theme 1	Theme 2	Theme 3	Theme 4	Theme 5	Theme 6
Responsible means patient-first, human-led, and role-based	Starting with contained workflows makes AI's value easier to prove	Context and role-based architecture beat raw model power	Governance must be an operating model, not an approval gate	Executives approve what they can compare	Workforce redesign is the real scaling bottleneck



Theme 1

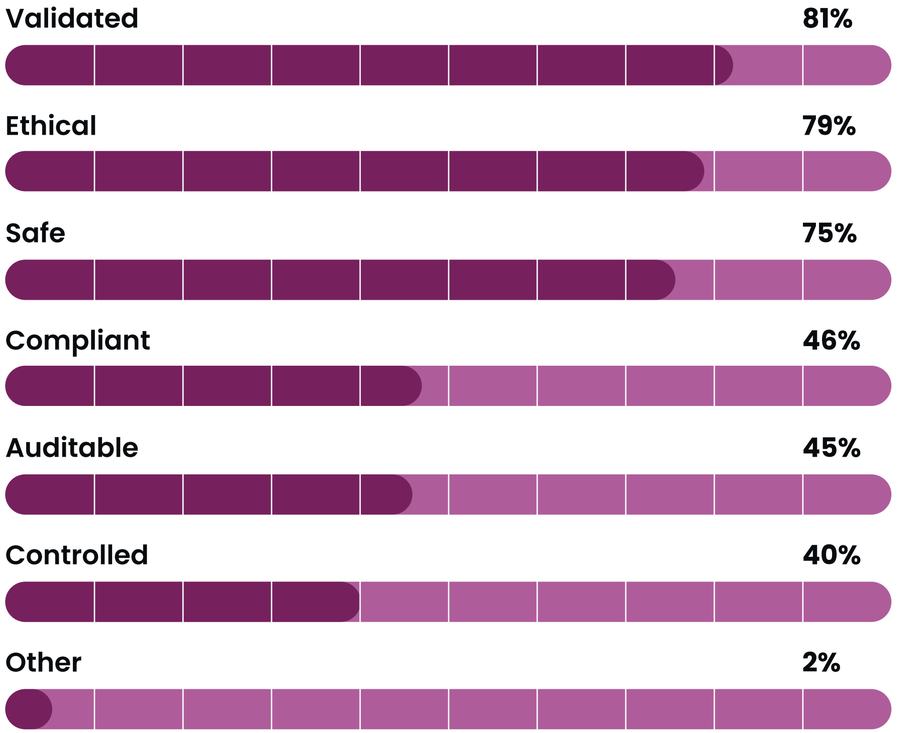
Responsible means patient-first, human-led, and role-based

Before anyone debated architecture, the roundtables kept returning to a single anchor: in Life Sciences, "responsible" is not a generic ethics label borrowed from tech. It is defined by who you ultimately serve. Participants grounded responsibility in patient welfare, accuracy, and accountability, and drew a hard line that humans, not systems, remain the accountable layer. This is why the most credible early designs are those that are role-based support systems that improve preparation, retrieval, and synthesis while leaving consequential action in human hands.

The implication runs through the rest of the report: responsible AI is a design constraint, not a disclaimer. Every scaled use case should start with who is being served, what decision is being supported, what harm is unacceptable, and when the system must stop and escalate.

From The Survey

What defines "responsible AI" in Life Sciences?



N = 135

Every attribute registered as relevant. Validated, ethical, and safe led by a wide margin: the vocabulary of an industry that answers to patients first.





London Roundtable

"We as doctors have a saying: primum non nocere — first, do no harm. My guiding principle would be, it's got to be about patients. It's got to help patients — and it's got to be able to understand whether there's potential for harming patients."



Philadelphia Roundtable

"Humans have to be accountable. Regardless of what the agents help us do — accelerate, project-manage, whatever — at the end of the day, we are accountable for the answers we provide. AI agents don't make this stuff up; they parse what comes out of humans. So unless you're willing to be accountable for that, you can't."



Basel Roundtable

"Being human-centric is actually an accelerant, not a decelerant."

Theme 2

Starting with contained workflows makes AI's value easier to prove

Asked where AI should start, leaders were strikingly consistent and strikingly unglamorous. Not autonomous decision-making — bounded, repetitive, high-friction work where the value is obvious and the blast radius is small: review and approval cycles, pre-call preparation, insight logging, triage, and coaching. Where the emphasis falls depends on where the work sits. Upstream commercialization roles, such as content development, analytics, and planning, sit further from the customer and can lead with efficiency: speed and volume. Downstream, customer-facing roles lead with effectiveness, for three reasons: they are closest to the customer, each interaction calls for more judgment, and the organization needs consistency and control across thousands of those interactions. Either way, leaders wanted proof: measure the current state, then measure again. The appetite was for compounding, measurable wins with AI, not moonshots.

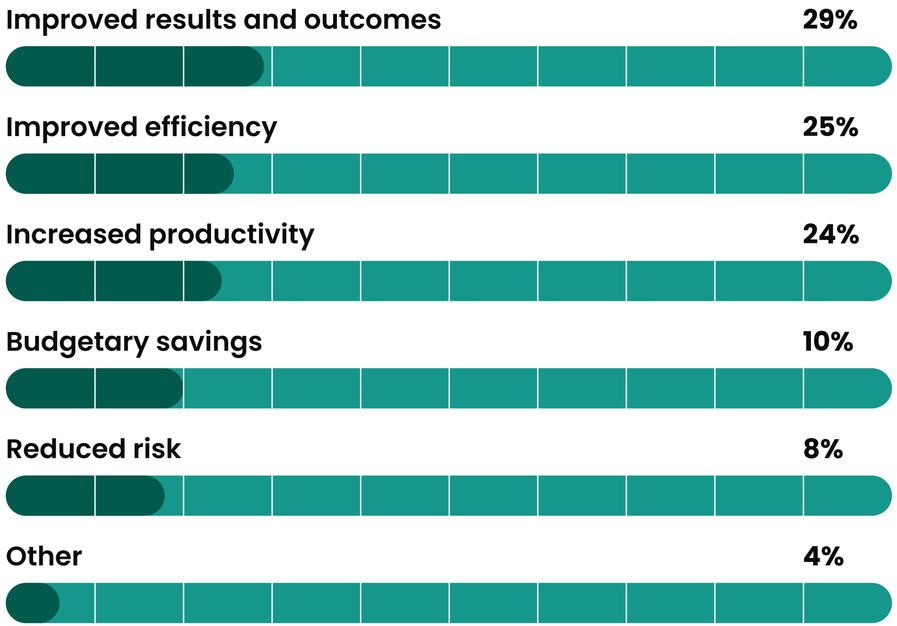
The pattern matters for sequencing. Start where value is high and governance complexity is low, then build toward harder use cases as AI trust and controls mature.





From The Survey

What is the most important benefit you expect from the use of AI at your organization?



N = 135

Better outcomes edged out efficiency and productivity as the benefit leaders want most. Budgetary savings trailed well behind.



London Roundtable

"For every initiative, I try to understand the current state — where you are now versus where you were. How much time do you spend doing this? Track it. This used to take 22 seconds, now it's four. I'm very deliberate, because that's the only way you keep gaining adoption: when you show value. Not in money yet, but in the time that's saved."



Philadelphia Roundtable

"You should not be playing in the efficiency pool. You should be playing in the effectiveness pool."



Toronto Roundtable

"[The tool] was untethered to the what-for — how's it gonna help me, help my job? No organization can move as fast as the external environment is moving. It's crawl, walk, run, fly."



Theme 3

Context and role-based architecture beat raw model power

A recurring correction to the hype was that the differentiator wasn't the size of the AI model, but the quality of the context it operated within. An agent that didn't understand the role it supported — the data it should draw on, the customers it serves, and the boundaries it must work within — wouldn't earn enough trust to be used, no matter how capable the underlying AI model. In a regulated industry like Life Sciences, context quality was also a governance issue. The right approved data sources, role boundaries, and escalation points were what made an agent safe, not just useful.

The design lesson was to build from the role outward, not from the model inward. The acceptable architecture for a field rep, an MSL, a key account manager, and a patient-support workflow would not be the same.

Online Survey Result



San Francisco Roundtable

"You're only as intelligent as what you're given. We were trying to use AI to understand doctors' prescribing behaviors. Where it falls short is if the information isn't getting fed. If doctors aren't reporting what conferences they're attending, the machine won't know it."



London Roundtable

"As an organization, I have a sales rep, a field reimbursement specialist, a medical science liaison, a medical information tech, and a nurse navigator [agent] I use to support customers across home, clinic, and institutional care. The magic is a complete understanding of the patient at the center, and who engages them."

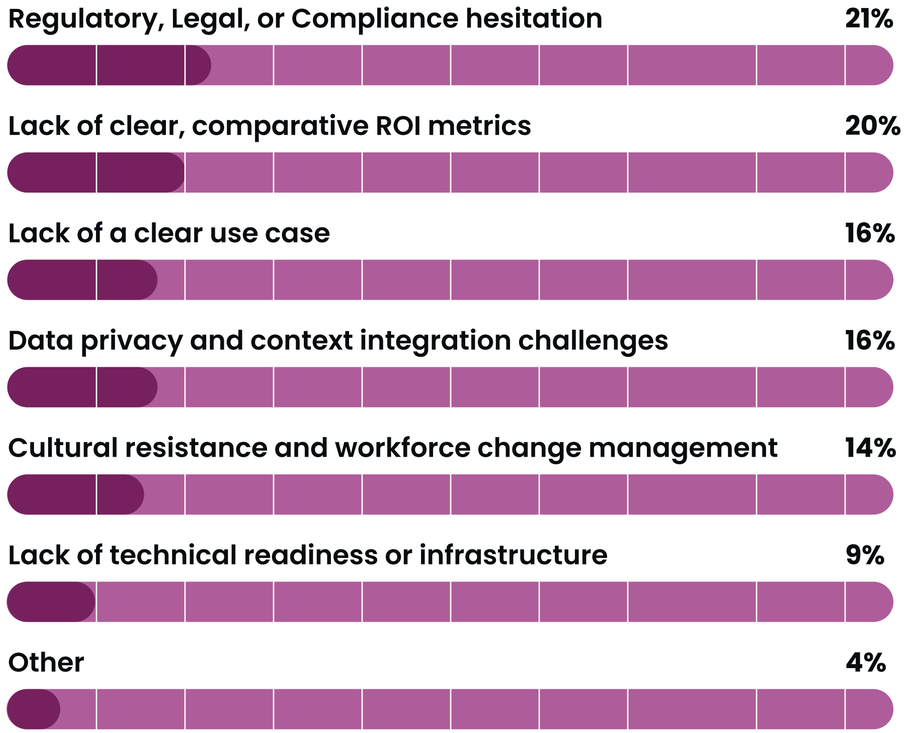


Princeton Roundtable

"For me, responsible means I've got to be absolutely sure I can give the right inputs and context for what I'm trying to discover, without that then being disclosed to the world. Without that context, it's not going to come up with the best answers."

From The Survey

What is the single biggest internal barrier to scaling AI within your organization?



N = 135

Data privacy and context integration tied for third. Technical readiness ranked near the bottom. The harder problem is feeding the system the right context, safely.



Theme 4

Governance must be an operating model, not an approval gate

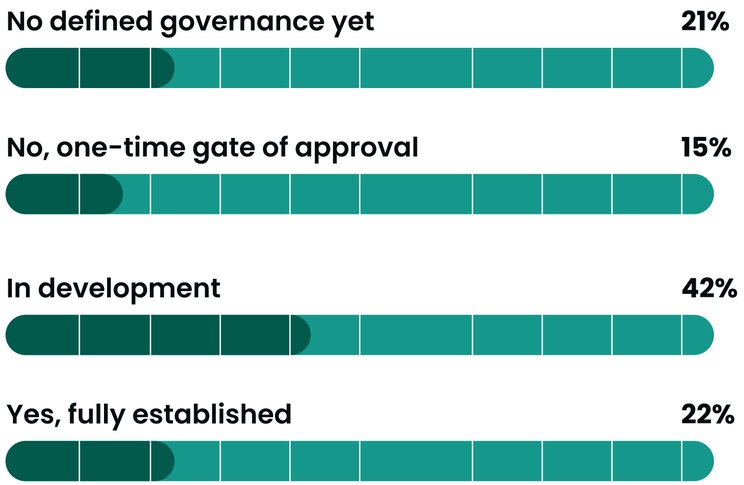
Governance was the most-cited concern when it comes to AI and, by the leaders' own admission, the most neglected in practice. The live danger came up directly: people inside the organization are already building AI agents with no governance around them — shadow AI is growing faster than oversight. The fix leaders described is not a one-time IT sign-off but a standing operating model: named business owners, clear escalation paths, audit trails, the ability to deploy and revoke an AI agent, and SOPs that apply at both the AI system and the sub-agent level. Legal was repeatedly cited not as a blocker but as an essential partner.

The Princeton roundtable's session with a 30-year Life Sciences lawyer sharpened the risk surface further. AI-recorded interactions are discoverable in litigation, adverse-event signals created trip-wire obligations, and disclaimers had to be enforced in the system. The throughline was clear. Governance is the difference between a system you can prove you control and one you merely hope is behaving.



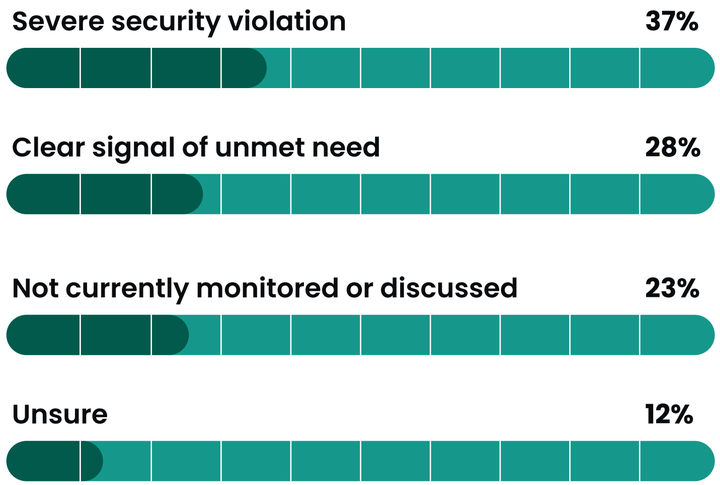
FROM THE SURVEY

Does your organization have an AI governance process (named owners, escalation paths, SOPs) rather than a one-time IT approval gate?

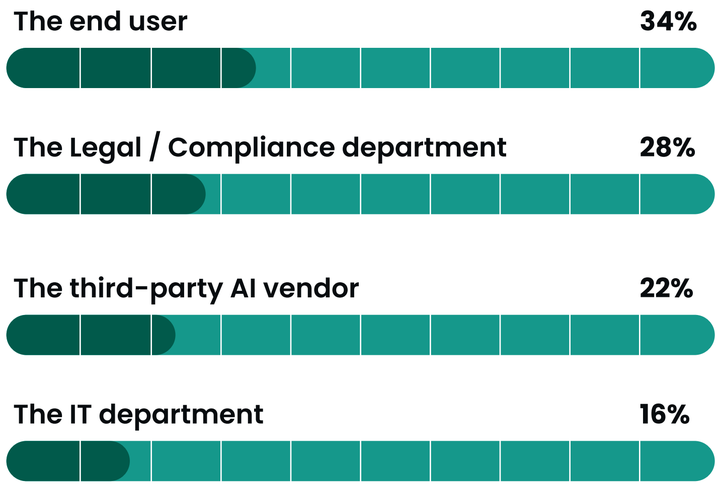


N = 135

How is employee use of unauthorized AI agents currently viewed in your organization?



Who should be held ultimately accountable for the actions an AI agent takes?



Only **22%** have governance fully established. Nearly a quarter do not monitor or discuss unauthorized agents at all. And no single owner commands a majority view on accountability.



Philadelphia Roundtable

"The standards are there; the AI ultimately has to do the same thing. Humans get things wrong sometimes too, but the decisions the AI is making have to be auditable. We can question them."



Basel Roundtable

"It's essential for each AI agent to know: when do I have to escalate to my human, and call him at the beach? We can't permanently be out at the beach; we need to come back and visit the accountability department."



San Francisco Roundtable

"MSLs themselves are building AI agents and starting to automate all that stuff, with no governance in place."



Theme 5

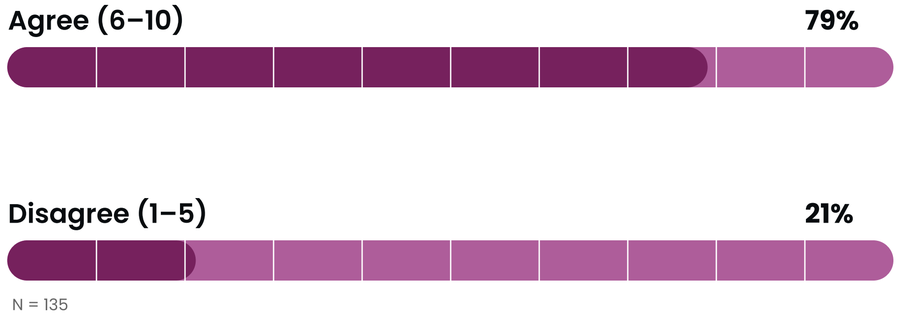
Executives approve what they can compare

The single loudest demand across the series was not for zero risk; it was for a fair comparison. Leaders pushed back hard on the idea that AI must be perfect; they wanted it to be measured against the current human baseline, which is itself non-zero given cognitive bias, fatigue, and memory lapses. The recurring request: give me a before-and-after on ROI, risk, quality, and adoption. And a sharp test for value was that an AI agent has to be measured against the old way of doing the work, not against perfection.

This is the demand the grading standard answers: a buyer or board needs an at-a-glance way to compare systems, and a function lead needs three numbers measured before and after an AI pilot. Comparison, made simple, is what unlocks funding.

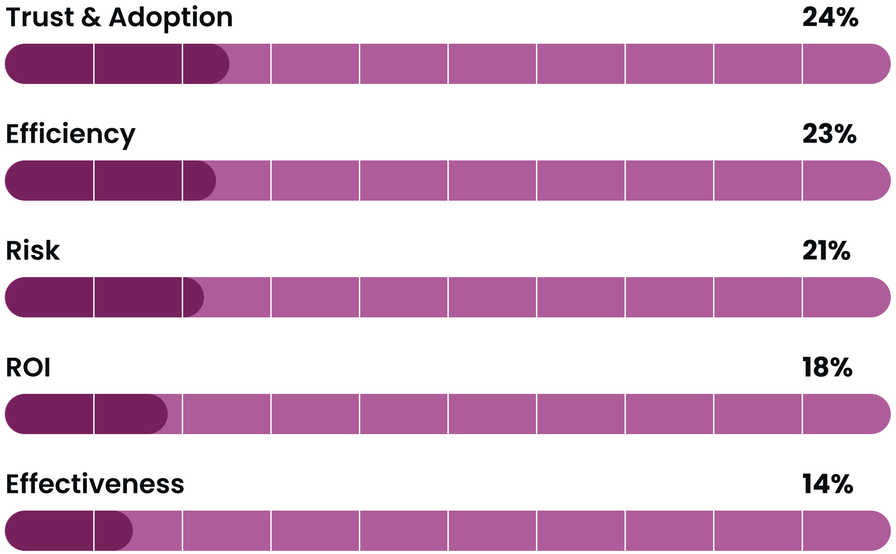
From The Survey

To what extent do you agree: "AI risk and ROI must be measured comparatively against the current human baseline, rather than zero-risk perfection"? (1 = do not agree, 10 = strongly agree)



N = 135

When evaluating an AI pilot's success, which measurement dimension is the most critical for your organization?



Four in five want AI judged against the human baseline. No single measure dominates how a pilot is judged, so a credible comparison has to hold several dimensions at once.



Basel Roundtable

"It's not that the risk doesn't exist in a human-only world. If your sales reps are detailing, they're very prone to cognitive error — but we've agreed how we quantify that risk. Give them a knowledge assessment, they score 80–90%; we've quantified it. That's where transparency becomes very important."



London Roundtable

"The underlying models are non-deterministic. It's not a perfect environment; it's prone to errors. The biggest risk I've seen is: how do you quantify that risk? That's the real limiter to pervasive adoption."



London Roundtable

"As we bring these thinking machines into organizations that shape the lives of our employees and patients — are we making right choices or weak choices? How do we know? What's that yardstick?"

Theme 6

Workforce redesign is the real scaling bottleneck

Roughly a third of every roundtable voiced that the hardest part of AI adoption isn't the technology; it's the people. Leaders were candid that cultural resistance and workforce change, not infrastructure, is what stalls scaling. The reframe that resonated across the board is that a job description should be a blueprint for the future, not a relic of the past. The work is to separate what the agent handles (the rote, data-processing tasks) from what the human owns (judgment, connection, strategy, the ethical override) and to say so honestly, because people adopt agentic AI faster when they know exactly how it changes their job. When they don’t, they just work around it.

Handled well, the anxiety about AI converts into excitement. People get time back for the work only they can do. Handled poorly, the best technology in the world dies in the org chart.





Toronto Roundtable

"Take a person's job description and put it in a centrifuge. What surfaces to the top? Data processing, rote tasks - this is what agents pick up. What settles at the bottom - human connection, judgment, strategy - is where humans lean in. That's your Future of Work Blueprint."



Boston Roundtable

"We all have to lean in and redefine the future of work for our teams... we have to intentionally spark these conversations internally."

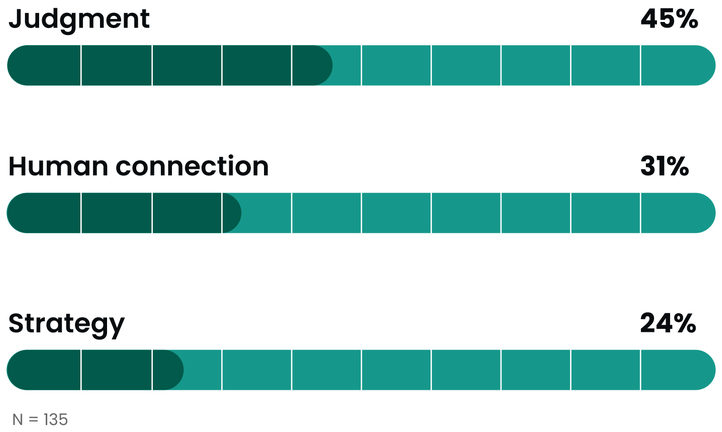


London Roundtable

"When I'm clear on what the role is, the job description is consistent with the expectation set for the human to do the job successfully. Agentic AI — agents, and ultimately super-agents — should complement that, and be held to the same standard as the individual."

From The Survey

Which skill do you think is most important for humans to retain?



N = 135

Judgment led clearly, followed by human connection and strategy: the same three skills a redesigned role should protect. Asked about barriers to scale, leaders also placed cultural resistance and workforce change **(14%)** ahead of technical readiness **(9%).**



SECTION 05

FROM THEMES TO A STANDARD: **THE RAIL SCORE**



Across the six themes, a clear picture emerged of the kind of AI system leaders would trust in the Life Sciences industry. It keeps a human accountable and central. It creates value you can point to. It works within the rules, can be governed and watched once it leaves the pilot, and keeps the organization's data and IP safe. Taken together, it can be trusted to protect the patient. This is the RAIL Score.

That gives us **six components of trustworthy agentic AI**, each tracing back to what RAIL roundtable and survey leaders raised:

Compliant	Capable
1. Governance 2. Security 3. Testing & Certification	4. Human Centricity 5. Value Creation 6. Orchestration

SIX COMPONENTS OF TRUSTWORTHY **AGENTIC AI**

Most of the market grades agentic AI on what is easiest to point at, Security and Governance, and stops there. Our market conversations point to four more components that must be graded to decide whether a system is trusted, used, and worth scaling: Testing & Certification, Human Centricity, Value Creation, and Orchestration. A system can be secure and well-governed and still be untested, unadopted, unmeasured, and stranded in a single tool. The RAIL Score grades all six components, and treats two of them, Security and Governance, as non-negotiable floors.

Compliant

Governance
SOPs, guardrails, and disclaimers enforced in the system, at both agent and sub-agent level, named owners, and escalation paths

Security
data integrity and IP protection; the floor everyone assumes, deploy and revoke

Testing & Certification
observability, black box to glass box, grade the agent's performance - certify the agent like a human

Capable

Human Centricity
built from the role outward; augments a person and preserves human ownership of consequential action, role-based context

Value Creation
measurable, comparative value against the human baseline

Orchestration
coordinates multiple agents, systems, and data sets to complete complex workflow end to end

HOW THE RAIL **SCORE WORKS**

Each of the six components is scored between **1 - 3** against a defined scoring rubric, and the six scores sum to a total out of **18**.

- **1 — Absent / ad hoc:** the quality isn't really there, or it lives outside the system in informal human effort.
- **2 — Partial / inconsistent:** present at the top level or in places, but uneven, manual, or untested.
- **3 — Established / enforced:** built in, consistent, and verifiable.

The RAIL Score descriptors describe the grade (1, 2, or 3) for each component.

⚑ **= floor component.** Security and Governance are safety-critical, so a strong total can't paper over a weak score on either of these components.

The Scoring Rubric & Descriptors

Component	1 - Absent/adhoc	2 - Partial/inconsistent	3 - Established/enforced
Human Centricity (role architecture · context grounding · human ownership of consequential action)	Generic, role-agnostic assistant; no human checkpoint on consequential action	Some role tailoring and context; human-in-the-loop exists but is informal or inconsistent	Built role-by-role from real workflows; rich job context; human ownership of consequential action explicit and enforced
Value Creation (value defined upfront · before/after baseline · ROI mapped to a number)	Value asserted, not defined or measured; no baseline, no metric	Value defined and a metric or two tracked, but no clean before/after comparison	Target value defined upfront, measured against the human baseline, ROI mapped to a real number leadership can compare
Orchestration (connected systems · agent-to-agent hand-offs · context carried across data sets, systems, and workflows)	System-bound AI working in isolation; agents cannot work across systems or with other agents, hand-offs to other tools are manual	Connects to some systems or agents, but hand-offs are brittle or context is lost between steps	Agents coordinate across systems, other agents, and approved data sets; context and guardrails are upheld at every hand-off
Governance ⚑ (named owner & escalation · in-system SOPs, guardrails & disclaimers · enforced at agent and sub-agent level)	No named owner; compliance lives in human review only; no in-system guardrails	Named owner; guardrails and disclaimers exist at the top level, but the sub-agent / skill level is uneven	Accountable named owner and escalation path; SOPs, guardrails, and disclaimers enforced in-system at both agent and sub-agent (skill) level
Security ⚑ (data protection · IP containment · access controls · deploy & revoke)	No clear data boundary; IP could leak to a model or downstream system; no one can say who can switch it off	Standard protections in place; gaps in IP containment or access control; revocation possible but manual and untested	Enforced data protection, IP contained, least-privilege access controls, tested deploy-and-revoke
Testing & Certification (observability · audit trail · performance graded & re-certified)	A black box: outputs are not logged or reviewed, and performance is never graded	Basic logging and spot checks; performance reviewed informally, with no re-certification schedule or trigger	A glass box: full observability and audit trail; performance graded against defined standards and re-certified on a schedule or based on a trigger, just like a human

TIERS: TURNING THE **SCORE INTO A DECISION**

The total RAIL Score maps to one of three actions:

TIER	SCORE	WHAT IT MEANS
High	15 - 18	Trustworthy enough for use in Life Sciences
Medium	10 - 14	Specific gaps to close before use in Life Sciences
Low	Below 10	Not fit for use in Life Sciences
Floor Rule	Security or Governance scored **below 3** caps the system at **Low**	

THE HARD FLOOR

Regardless of total, the two safety-critical components gate the RAIL Score result: an AI system or tool can't be rated High if it falls short on Security or Governance. This is what stops an AI system or tool from buying its way to "High" on strength elsewhere, while simultaneously introducing unacceptable risk to the company.



RAIL SELF-GRADING SCORECARD

Use the RAIL Scorecard

Download now

Assess an AI tool your organization uses, is piloting, or is considering using the RAIL Score. It will help your team reach a shared verdict and identify any gaps that need attention.

WHAT A RAIL SCORE RESULT **LOOKS LIKE**

The RAIL Score is built to show which dimension is dragging, not just hand down a verdict.

Worked Example

System: MSL pre-call prep AI agent
An agent that assembles briefing material before a medical science liaison meets a healthcare provider.

TOTAL RAIL SCORE:

14/18

TIER:

Medium. Both floor components are met

Recommended next steps:

- Define one comparative prep-time number (before/after)
- Grade a sample of briefs for accuracy and set a re-certification date or trigger.
- Close the CRM hand-off so call notes flow back without manual re-entry.

COMPONENT	SCORE	WHY
Human Centric	03	The AI agent is built for a specific role (MSL) and the specific workflows required of that role, with the MSL reviewing every pre-call brief before use.
Value Creation	01	Everyone believes the AI agent saves time, but no one has measured prep time before vs. after, so impact is missing.
Orchestration	02	The AI agent draws on the CRM and the approved medical content library, but the MSL still copies notes back into the CRM by hand.
Governance	03	There is a named medical affairs owner of the AI agent; guardrails and disclaimers are established and enforced in-system, down to each sub-agent.
Security	03	The AI agent runs inside the company environment; no data leaves to an external model. The owner can switch it off if security is at risk, and has tested doing so.
Testing & Certification	02	The outputs of the AI agent are logged, but no one has graded the briefs for accuracy or set a re-certification date or trigger to ensure the AI agent is still performing at the required level.



SECTION 06

THE AGENTIC AI PLAYBOOK: **WHERE TO START, WHAT TO BUILD**

The RAIL Score tells you whether an AI system or tool is trustworthy. It doesn't tell you where to point AI in the first place. And most teams have more ideas than they can sensibly take on at once. So the question becomes: of all the places you could use AI in commercialization, where do you apply it first?

WHERE TO START?

The leaders we spoke to who were furthest along with AI answered the question the same way. They weighed each idea on two simple questions: how much value would it create, and how hard would it be to govern it safely? That sorts every use case into one of three groups.

Start now:
high value, easy to govern.

The safe, obvious wins: review and approval cycles, pre-call prep, insight logging, triage, coaching. These are where most teams should begin.

Build next:
high value, but harder to govern.

Worth doing once the controls are in place: connecting systems together, supporting a full product launch, anything that chains several agents together.

Defer or constrain:
high risk.

Anything that could reach a patient or regulator without a human in between: autonomous external messaging, direct patient advice. Not never; just not first.

Whichever you pick first, decide up front how you'll know it worked. It is difficult to build a business case just on projected savings, so choose a few simple measures, take a reading before the pilot, and take it again after. A real before-and-after on the actual work.

MATCH THE APPROACH **TO THE ROLE**

Where a role sits in commercialization shapes what an AI agent should do for it. Roles differ in how much judgment each task demands, and in how much consistency the organization needs across every interaction.

	UPSTREAM ROLES	**DOWNSTREAM ROLES**
Examples	Content development, analytics, forecasting, planning	Field reps, MSLs, market access, patient services
Distance from the customer	Further away; work feeds other teams	Closest; work reaches HCPs and patients directly
Judgment per task	Lower; much of the work repeats	Higher; every conversation is different, understanding boundaries is critical
Lead with	Efficiency: more output, faster	Effectiveness: better-prepared interactions, with consistency and control at scale
Where AI agents start	Taking volume off the team	Preparing and supporting the human, who still makes the call

Whichever role you pick first, decide up front how you'll know it worked. It is difficult to build a business case just on projected savings, so choose a few simple measures, take a reading before the pilot, and take it again after. A real before-and-after on the actual work will ensure true value is understood.

WHAT TO BUILD?

Giving the AI agent a job description

Once you know what to build first, the next move is to scope it properly, and articulate, in plain terms, what the AI agent is for. This is, in effect, a job description for the agent, the same discipline a team would apply to any new hire.

Before committing to an AI agent, a team should be able to answer:

1. **What role does it support?** Identify which human role the agent supports.
2. **What is it responsible for?** The specific task or workflow it owns, and the outcome it's meant to produce.
3. **Who owns it?** The named human accountable for it, and who it escalates to.
4. **What can it do on its own, and what can't it do?** Where its authority ends and a human steps in.
5. **What data and sources can it use?** The approved inputs it draws on, and the boundaries it can't cross.
6. **How will we know it worked?** The few before-and-after measures that prove its value.



AGENT JD TEMPLATE

Build Your Agent's Job Description

Download now

Use this template to turn the six questions above into a one-page AI agent brief that business, IT, and compliance can agree on before work begins.



SECTION 07

REDESIGNING WORK:
THE HUMAN-CENTRIC MANDATE

Every other section in this report is about the technology. This one is about the people who have to work alongside it. When we asked what stalls AI in practice, the answer was rarely infrastructure or budget. It was the workforce: the subliminal resistance of people who don't know what the technology means for their own job.

The fix for this isn’t a reassuring All-Hands Meeting. People fear ambiguity about their place in the new setup, and specificity is what counters that. The moment a role is drawn clearly, showing which work the AI agent supports and augments, and what the person is now freed to own, most of that worry drains out of the room. So the work for leaders is to redraw the job, honestly and in detail.

Redrawing a role comes down to a single move: separate what the AI agent does from what the human keeps. The AI agent takes the rote, repeatable work, retrieval, formatting, first-draft assembly, logging. The human keeps what only a human can hold: judgment, human connection, strategy, and the final call when something doesn't look right. Done well, this becomes an adoption accelerant, because people can see what they gain by having an AI agent “helper.”

STORY FROM **THE FIELD**

In a rare-disease patient-support team, one specialist's day was being eaten by retrieval, readying records and paperwork before each patient interaction.

Hand that retrieval to an agent and the role doesn't shrink, it shifts. The agent does the pulling-in; the person gets the time back for the patient.



Boston Roundtable

"I can take someone who's trained as an NP or social worker, and let them be all about the passion they have, and serve patients instead of pulling in the form. If I can keep the patient service person focused on the patient, and not on the screen, they can take on additional therapy and indication."

JOB RE-DESIGN WORKSHEET

Redraw Roles Around the Work

Download now

Use this worksheet to separate what the AI agent handles from what the human owns, define the hand-offs between them, and decide how you will measure whether the redesign is working.

MANAGER'S TEAM CONVERSATION SCRIPT

Bring Your Team Into the AI conversation

Download now

Use this talk-track to explain how an AI agent will change the team's work, what remains firmly in human hands, and how it'll benefit the team.



SECTION 08

THE FIRST **90 DAYS**

The hardest part of any framework is the first move. This section turns everything from the previous sections into a single quarter of concrete action. Set up to build momentum, prove value early, and give a team something real to show before the next planning cycle.

Days 0–30

Take stock and draw the lines.

Inventory what's already running with the AI Inventory tool.

Stand up a small, cross-functional RAIL Council: commercial, medical, IT, legal. Agree your red lines: the uses you won't allow yet. And pick one AI tool or system to put through the RAIL Scorecard.

Days 30–60

Grade one, launch one.

Score that first system with the RAIL Self-Grading Score card and see where it lands: High, Medium, or Low. Then start one pilot, scoped with an AI Agent JD template, so it has an owner and clear boundaries from day one.

Lock the few before-and-after metrics by which you'll judge it on before it goes live; not after.

Days 60–90

Measure and decide.

Re-measure against the baseline you set. Run the Manager's Team Conversation Script with the team whose work the agent touches. For specific individuals or teams, share the Job Redesign Worksheet so the people side moves with the technology. Then make the call the scorecard points to: scale it, fix the specific gaps, or stop.

CHECKLIST

90-Day Launch Checklist

Download now

Get a working copy of every step above as a tickable list, with the tool to use at each stage, so your team can track its first quarter from inventory to decision.

SECTION 09

GRADING A VENDOR **BEFORE YOU BUY**



Most agentic AI solutions enter a Life Sciences organization through procurement. A demo shows an AI tool or system at its best. The RAIL Score asks how it behaves on an ordinary day: with real data, real users, and a regulator who may one day ask for the record. Buyers carry the risk in the gap between the two.

The **Vendor-Grading Checklist** turns each of the six RAIL Score components into direct questions a buying team can put to any vendor, in an RFP, a demo, or a security review.



CHECKLIST

Vendor-Grading Checklist

Download now

Put evidence-based questions to any agentic AI vendor across the six RAIL components, and score the answers before procurement moves forward.

RAiL

Responsible AI
in Life Sciences

REPORT 2026:

Defining an Agentic AI Trust Benchmark

ACTO

