WAVESTONE

The AI Value Reckoning in Life Sciences

Why the Era of Unbridled AI Spending is Ending in Biopharma

The Era of Unbridled AI Spending is Ending for Biopharma

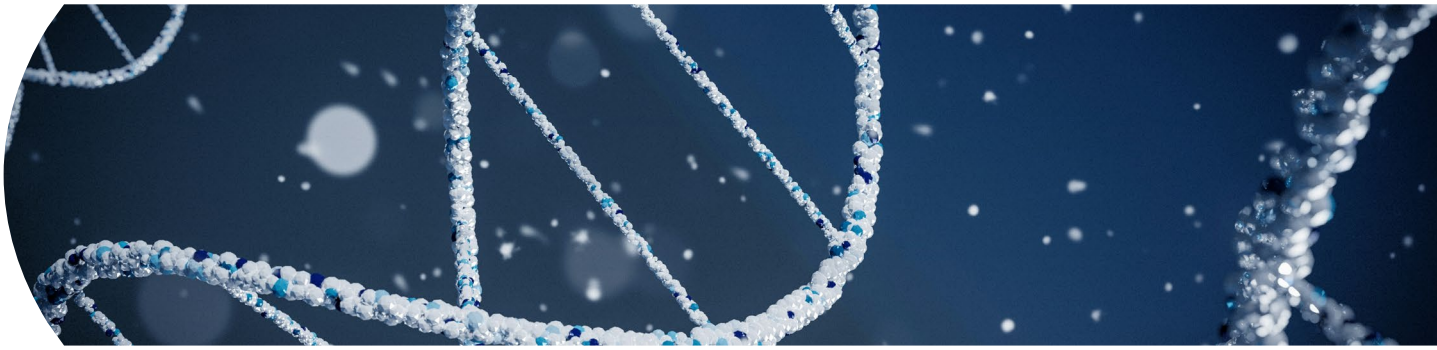
Life sciences companies invested early and aggressively in AI. Discovery teams deployed models to accelerate target identification, clinical groups experimented with protocol optimization, safety functions tested enhanced signal detection, and commercial teams piloted AI-assisted HCP engagement and medical writing. Early results looked promising, but the targeted value has been slow to materialize at enterprise scale.

Declining R&D productivity, lengthening development timelines, a historic patent cliff putting over $300 billion in sales at risk through 2030 and increased regulatory scrutiny are intensifying the pressure on leaders to deliver value.

- Across all industries, **80% of CEOs** believe their job is at risk if AI fails to deliver this year.
- **81% of U.S. CEOs** believe a peer would be removed over an unsuccessful AI strategy.
- S&P 500 disclosures of board-level AI oversight have increased **84% year-over-year and 150% since 2022**.
- Shareholder proposals related to AI governance have **quadrupled**.
- A May 2026 Dataiku / Harris Poll found that **65% of CEOs** are now more worried about over-investing than about falling behind.

Source: Dataiku / Harris Poll CEO Survey (2026); ISS Corporate analysis of S&P 500 filings (2025)

The Reckoning Has Begun



Across all industries, eighty percent of CEOs believe their job is at risk if AI fails to deliver this year. Eighty-one percent of U.S. CEOs believe a peer would be removed over an unsuccessful AI strategy. S&P 500 disclosures of board-level AI oversight have increased 84% year-over-year and 150% since 2022. Shareholder proposals related to AI governance have quadrupled. A May 2026 Dataiku/Harris Poll found that 65% of CEOs are now more worried about over-investing than about falling behind.

The Starting Point

In the late 1990s and early 2000s, enterprises poured billions into ERP implementations on the promise of operational transformation. By year three of those programs, a wave of CIO terminations swept through the Fortune 500. Boards gave the technology two budget cycles to prove itself, watched integration costs run past the original business case, and replaced the executive who owned the program. Published failure rates from that era ranged from 55% to 75%, and the average implementation took three to five years to reach measurable value, if it reached it at all. We are now eighteen to twenty-four months into the enterprise AI investment cycle, and the early signals are tracking the same curve.

The Curve Is Steeper in Life Sciences

46%
Lack a structured framework to measure AI ROI
Wavestone, 2025

95%
of GenAI pilots produced zero measurable P&L impact
MIT NANDA Initiative, 2025

56%
of CEOs report no revenue or cost benefits from AI
PwC CEO Survey, 2026

16%
of AI initiatives have scaled enterprise-wide
IBM CEO Study, 2025

20%
are achieving revenue growth from AI despite 74% having aspirations to do so
Deloitte, 2026

79%
of life sciences organizations slowed AI deployments due to governance and compliance issues
Modus Create, 2026

70-80%
of AI pilots sciences fail to reach controlled production
USDM Life Sciences, 2026

In life sciences, the curve is steeper. The top sixteen pharmaceutical companies spent $159 billion on R&D in 2025, down 3.6% from the prior year, as pipeline prioritizations forced difficult choices about where investment continues and where it stops. Biopharma's internal rate of return on R&D investment has fallen to 4.1%, well below the cost of capital, while Phase I success rates have dropped to 6.7%, down from 10% a decade ago. ZS's 2026 CDIO research found that nine in ten pharma technology leaders now see growth pressures from competitors' scientific advancements to AI disruption and regulatory friction as active threats to business growth. The question is no longer "where can AI work?" but "where will AI drive the growth that matters?"

The patience that funded the experimentation phase is gone. Boards and C-suites have shifted from asking how to get started to asking where the return is, and the gap between investment and answer is widening.

A December 2025 Teneo study quantifies the expectations problem: 53% of institutional investors expect AI ROI in six months or less, but only 16% of large-cap CEOs think that timeline is realistic.

Wavestone's 2025 Global AI Survey of 500 executives across six countries captured the operational reality underneath these board-level pressures. Ninety-nine percent of respondents have deployed at least one AI solution. Seventy percent place AI at the heart of their strategy. And 46% still lack a structured framework to measure whether any of them are working. Deployment is largely settled. The value question is not.

The AI value reckoning has started, and most enterprises are at risk of not passing it.

The organizations that will are making three design choices the majority systematically deferred. They are designing for trust, for accountability, and for capacity reinvestment. These are operating-layer decisions that determine whether AI investment compounds into enterprise capability or evaporates into activity. That is the work the strategy decks tend to leave out.

The Five Deferred Design Decisions

The gap between deployment and value will not be resolved through patience alone. It is structural.

Organizations producing value made operating decisions the rest of the market deferred. They decided how work changes when AI enters it, how data flows when models need it, how governance operates while systems are being built, how outcomes are are being built, how outcomes are measured, and how leadership stays engaged through early failure.

Five deferred design decisions now explain where value is leaking.

1. **Scaling pilot deployments without redesigning the work**
2. **Waiting for clean data instead of architecting for access**
3. **Building AI on operating environments that are not assured**
4. **Measuring deployment, not behavior or business outcomes**
5. **Losing sponsor continuity through the first failure**

1. Scaling pilot deployments without redesigning the work

Pilot environments produce results because a small team adapts. Enterprise environments need the work itself to change, and many have not redesigned it. Tools go live, dashboards register adoption, and the underlying work looks the same as it did before deployment. Wavestone's 2025 Global AI Survey captures the operational consequence: only 30% of target users have meaningfully changed how they work, even in organizations reporting broad deployment. Only 7% report that more than half their target users have significantly shifted day-to-day practices. Deloitte's 2026 research found that 84% of companies have not redesigned jobs around AI. Behaviors revert under pressure. The business case fails to materialize without anyone noticing. Boards see the spend without seeing the work change.

In life sciences, this pattern is especially visible. NVIDIA's 2026 State of AI in Healthcare and Life Sciences survey found that 70% of organizations are actively using AI, up from 63% in 2024, and 69% are using generative AI and large language models. Yet adoption and workflow change are not the same thing. A discovery team using AI-assisted target identification, a clinical operations group piloting protocol optimization, a safety function experimenting with signal detection may all show promising results within a contained environment, but unless trial design processes, cross-functional handoffs, and downstream regulatory workflows are redesigned, those gains do not translate into shorter development timelines or lower cost per asset.

As one clinical AI strategy lead noted in the NVIDIA survey, "The organizations seeing impact are those that embed AI into existing workflows instead of layering AI on top as a separate tool."

2. Waiting for clean data instead of architecting for access

The instinct to halt AI work until the data is fixed has stalled more programs than it has saved. Stanford's research found that only 6% of successful AI implementations had fully clean data. In 88% of cases, large language models unlocked previously inaccessible data sources: voice transcripts, scanned documents, regulatory submissions, clinical documentation, claims notes, legacy code, scattered knowledge bases. The organizations producing value architected for access rather than waiting for perfection. The enterprises still waiting are watching that advantage accrue to competitors who started without it.

Life sciences organizations sit on some of the most complex and fragmented data environments in any industry. Clinical study reports, investigator notes, adverse event narratives, medical literature, regulatory submission archives, and real-world evidence from electronic health records and claims databases exist across dozens of systems and formats. The instinct to wait for harmonization before deploying AI is particularly strong here, and particularly costly. The organizations producing value are using large language models to unlock these previously inaccessible sources, extracting structured insight from unstructured safety narratives, accelerating literature review across therapeutic areas, and enabling cross-study analysis that manual processes could never support at scale.

3. Building AI on operating environments that are not assured

Thirty-five percent of all project drag in enterprise AI programs comes from staff functions — Legal, Risk, HR, and Compliance — outpacing end-user resistance by a wide margin. Governance built downstream of delivery cannot govern delivery. When the teams building and operating AI are also the teams expected to govern it, the governance function becomes a rubber stamp or a bottleneck, and neither path leads to scale. Independence has to be designed in. Few enterprises have.

In life sciences, this failure is amplified by regulatory structure. AI systems that intersect with GxP processes face validation requirements, audit expectations, and documentation standards that extend well beyond initial deployment. The FDA's 2025 draft guidance introduced a risk-based framework for assessing AI model credibility in drug submissions. The EU AI Act classified certain life sciences AI applications as "high-risk," requiring conformity assessments and post-market monitoring. ISPE published its first GAMP Guide for Artificial Intelligence in July 2025, establishing lifecycle governance expectations for AI in regulated environments. Modus Create's 2026 research found that 79% of life sciences organizations slowed AI deployments specifically due to governance and compliance issues. Governance in this industry is not an internal preference. It is a regulatory requirement. When it is absent, the consequence moves beyond delay to exposure.

4. Measuring deployment, not behavior or business outcomes

Measuring deployment, not behavior or business outcome. License counts, model-in-production tallies, and pilot success rates have become the default proxies for AI value. They reveal nothing about whether work has changed or the P&L has moved. Deployment is a milestone, not a value indicator. Wavestone's 46% figure on missing measurement frameworks captures the consequence at scale, and Deloitte's data captures the cost: 74% of organizations aspire to revenue growth from AI, only 20% are achieving it. The measurement gap is a design failure that begins before deployment, when business outcome KPIs are deferred in favor of activity metrics that are easier to produce. By the time the board asks for evidence of value, the instrumentation to provide it does not exist.

In life sciences, value is measured at the level of the pipeline and the product: time to IND filing, enrollment velocity, probability of technical success, regulatory cycle times, launch uptake, cost per asset. These are the metrics that boards and investors care about. Yet most AI measurement frameworks in the industry track activity — number of pilots running, models deployed, users enabled — rather than whether any of those pilots shortened a development timeline or improved a submission outcome. The disconnect is particularly costly in an industry where R&D margins are projected to decline from 29% to 21% of revenue by the end of the decade and where every month of delay in a clinical program carries direct economic consequence.

5. Losing sponsorship continuity through the first failure

Losing sponsor continuity through the first failure.

Sixty-one percent of successful enterprise AI projects, according to recent research, included at least one prior failure. In every trackable case, the same executive who oversaw the failed attempt led the successful one. Sponsor continuity through the first stumble was a consistent differentiator, along with the organizational permission to fail that came with it. The SAP era taught the same lesson and the enterprise market forgot it. Programs that survive their first failure carry that learning into the second attempt. Programs that lose their sponsor restart the political clock, restart the budget conversation, and erase the organizational learning that was the most valuable output of the first attempt.

These five failures compound. An enterprise that deploys without workflow redesign, waits for clean data, governs after the fact, measures deployment instead of outcomes, and replaces its sponsor at the first stumble will not recover by buying more AI. It will recover by rebuilding the operating layer underneath the AI it has already bought. That rebuild is overdue across most of the Fortune 500, and the people setting the timeline are no longer interested in waiting.

Agentic AI makes the deferred decisions more expensive. Autonomous systems amplify every one of the five failures and remove the slack that has allowed enterprises to defer the design work. Workflow redesign cannot wait when agents are taking actions humans previously took. Data access becomes structurally consequential when agents are reasoning across systems. Governance built after deployment is governance built after the system has already acted. Deployment metrics actively mislead when the work an agent does is harder to observe than the work a human does. Sponsor continuity becomes more fragile, because the political consequence of an agentic failure lands harder than a recommendation-engine failure. Enterprises that have not solved these design failures at the generative AI layer are about to attempt agentic AI on the same foundation. The reckoning will arrive there faster.

In life sciences, agentic AI is already moving from concept to pilot. NVIDIA's 2026 survey found that 47% of healthcare and life sciences organizations are using or assessing agentic AI, with early applications in knowledge retrieval, research paper analysis, and clinical workflow automation. The governance gap is particularly dangerous here: an autonomous agent acting on safety data, regulatory content, or clinical trial parameters without adequate boundaries, escalation thresholds, and audit trails introduces risk that regulators are already watching for. Enterprises that have not resolved the design failures at the generative AI layer are about to compound them at the agentic layer, in an industry where the consequences of uncontrolled AI actions carry regulatory, legal, and patient safety implications.

What the Organizations Beating the Cycle Are Doing Differently

The enterprises producing measurable AI value are running ahead on three design decisions the rest of the market has deferred. They are designing for trust, designing for accountability, and designing for capacity reinvestment. These are decisions about how the enterprise operates around AI, not about the AI itself. They are also the decisions that determine whether AI investment compounds into capability or evaporates into activity.

Design for Trust

Trust means making the AI system's contribution defensible at the point where a professional puts their name to the work.

Design for Accountability

Accountability means making explicit decisions the market has been leaving implicit.

Design for Capacity Reinvestment

AI does not create business value. Leaders do. Where the recovered capacity gets redirected, and against what strategic objective.

Design for trust

Generative AI changed the adoption question, and agentic AI changed it again. Earlier waves of enterprise technology asked people to learn new tools and new ways of working. AI asks them to incorporate the tool's reasoning, judgment, and increasingly its actions into their own work product and stand behind the result. That is a structurally different request, and it makes trust the precondition for value.

A clinical scientist incorporating AI-assisted analysis into trial design, a regulatory affairs specialist submitting documentation supported by AI-generated summaries, a pharmacovigilance analyst interpreting AI-flagged safety signals, a medical affairs lead shaping scientific narratives with AI copilots are each doing something earlier technology did not require of them. They are extending their accountability to include the machine's contribution. Adoption requires that the extension feel defensible. If it does not, the tool is used minimally, worked around, or overridden.

The signals are already visible. Override rates in clinical decision support systems frequently exceed 40% even when technical performance is acknowledged as strong. The same pattern surfaces across knowledge-work domains as generative AI moves from experimentation into the parts of the job where professional judgment lives. Performance is necessary. It is not sufficient.

In life sciences, trust carries structural weight that other industries do not face. Clinicians remain accountable for patient outcomes. Scientists must be able to defend results under regulatory scrutiny, sometimes years after a submission. Regulators expect clear documentation, governance, and monitoring that extends well beyond initial deployment. These are not cultural preferences. They are professional and legal obligations. A regulatory affairs organization at a global pharma client illustrated this clearly when they piloted AI-supported translation for regulatory submissions. The model performed well during the pilot — high accuracy, substantially faster output. But upon rollout, leadership did not immediately discontinue outsourced human translators. Regulatory submissions are examined in detail and may be revisited years later; errors carry implications across jurisdictions. The team ran both processes in parallel, comparing outputs, identifying discrepancies, and building familiarity with edge cases. It looked inefficient. Costs overlapped for longer than anticipated. But as professionals saw quality and consistency for themselves, confidence grew, and reliance on human translators decreased without mandate. Adoption became stable rather than fragile. That team ultimately saved over $2 million per year.

Designing for trust means making the machine's contribution defensible at the point where a professional puts their name to the work. That has structural requirements: clarity about how a recommendation was generated and where its limits lie; reproducibility and traceability that hold up under later scrutiny; documentation, governance, and independent assurance built in from deployment rather than retrofitted after the first audit.

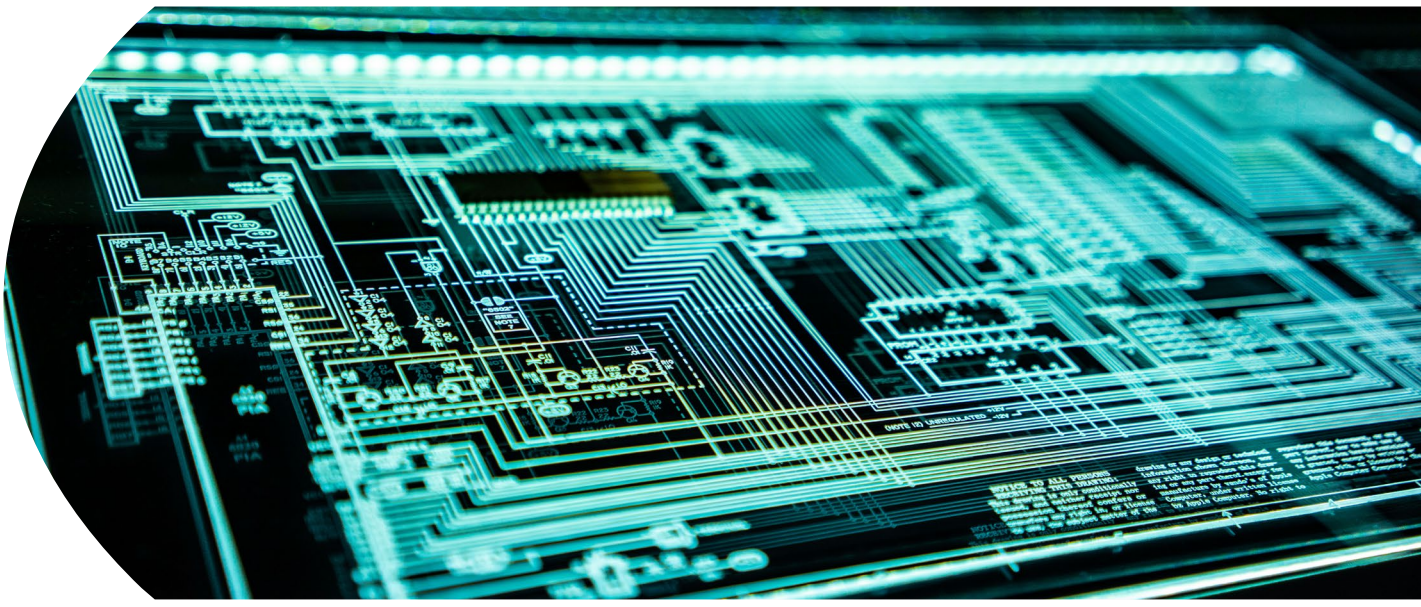
Design for accountability

In most large organizations, AI tools addressing similar problems are running in parallel, aligned locally, few connected through shared governance or performance metrics. Organizations with clearly defined digital operating models for AI are nearly twice as likely to realize measurable value as those operating with fragmented governance. When ownership is unclear, scale is not delayed. It becomes mathematically unlikely.

Designing for accountability means making explicit decisions the market has been leaving implicit. Where do AI investments report? Who approves the autonomy boundaries on agentic systems? Where does the escalation path go when an agent acts outside expected parameters? How are performance metrics reconciled across functions whose incentives may not align? Which staff functions have decision rights on what, and when in the lifecycle? These are operating model questions, and the strategy decks rarely answer them. The reckoning lands on the enterprises that did not.

In life sciences, accountability questions carry regulatory weight. When AI informs a labeling decision, a safety assessment, a clinical endpoint analysis, or a manufacturing quality judgment, the question of who is accountable is not organizational — it is legal. The FDA's enforcement posture has made this explicit: if AI influences GxP decisions, the entire solution must meet device-level quality, validation, and lifecycle controls. Accountability cannot be distributed across a model designer, a data engineer, a supervising scientist, and an executive sponsor without someone owning the outcome. Staff functions in Quality, Regulatory, and Compliance intuit this exposure correctly, which is why their resistance rises as AI ambition rises. Designing accountability before scaling is the only path through that resistance.

Agentic AI sharpens this further. When an agent misfires, accountability fragments. The agent designer, the orchestration framework, the data the agent acted on, the supervising human who did not intervene, the executive who approved the autonomy boundary — each of these can become the locus of responsibility, which usually means no one does. Staff functions intuit this exposure correctly, which is why their resistance is rising as agentic ambition rises. Designing for accountability before scaling agentic AI is the only way that resistance gets resolved.



Trust is an architectural decision, not a change-management afterthought. Enterprises that treat it as the latter watch adoption stall in exactly the places where the business case lives. Designing for trust feels slower at the outset. It accelerates time to business value.

Agentic AI raises the stakes on every part of this. Autonomous systems take actions before review rather than producing recommendations subject to it. The professional incorporating that action into their work product is doing so after the fact. Override is not available when the action has already happened. The trust imperative under agentic shifts to decision boundaries, escalation thresholds, transparent action logs, and human-on-the-loop design where the stakes warrant it. Enterprises that have designed for trust at the generative layer have a foundation to build on. Enterprises that have not are about to attempt agentic deployment without one.

Design for capacity reinvestment

AI does not create business value. Leaders do.

A successfully implemented, integrated, and adopted AI capability produces capacity in three forms. Time recovered on existing work. Work eliminated entirely. Decisions accelerated through better support. Each form is potential, not value. Whether the potential converts depends on a leadership decision most enterprises have not made: where the recovered capacity gets redirected, and against what strategic objective.

Fewer than a quarter of enterprise AI initiatives produce measurable financial impact, and more than 60% of executives report difficulty demonstrating bottom-line results beyond specific pilot use cases. The technology is working, yet the capacity it produces is not being deliberately redirected.

The default outcome of unredirected capacity is absorption. When an analyst saves four hours a week, those hours fill with adjacent work that was already in the queue. When a function eliminates a recurring task, headcount stays the same and other activities expand to fill the available time. When a decision cycle shortens, the slack disappears into review cycles, internal meetings, or workload that had been deferred. Absorption requires no decision. It is what happens when leadership does not act. Redirection requires an explicit decision about what strategic objective the recovered capacity will serve, who owns the redirection, and how the impact will be measured.

In life sciences, unredirected capacity has a direct and quantifiable cost. If a review cycle in regulatory affairs shortens by 20%, what changes? If signal detection becomes more efficient, how is that capacity redirected? If trial design accelerates, what strategic objective does that enable? When medical writing tools reduce drafting time for clinical study reports but the submission timeline does not change, the capacity was absorbed. When AI-assisted patient matching improves enrollment velocity but the freed clinical operations capacity is not redeployed against the next bottleneck, the value never reaches the pipeline. In an industry where every month of delay in a Phase III program can represent tens of millions in lost revenue, the cost of absorption is not abstract. It is the difference between a first-mover advantage and a competitive loss.

Most leadership systems are not designed to make those decisions at the speed AI is producing capacity.

Capacity therefore defaults to absorption, and the value never appears on the P&L. The enterprises producing measurable financial impact made the redirection decisions before deployment, not after. They knew which strategic objectives the AI investment was funding, in concrete terms, before the technology went live.

This is the layer where outcome measurement, workflow redesign, sponsor continuity, and use case portfolio strategy converge. It is also the layer where agentic AI changes the math most dramatically. Agentic systems do not just speed up tasks, but they remove them entirely. Recent research found that agentic implementations delivered median productivity gains of 71%, against 40% for high-automation approaches that retain task structure. Enterprises that have not decided where that capacity is going are about to deploy systems that produce it at scale.

These three design decisions are not parallel options. They are sequenced. Trust enables adoption. Accountability enables scale. Capacity reinvestment enables value. An enterprise can deploy AI without making these choices. It cannot capture value without making them. The organizations that will pass the reckoning made the choices early. The organizations that will not are still treating them as deferred work.

Cecilia Edwards,
Partner

Where to Start

Book an executive briefing with our team

Discuss the key takeaways from the paper, pressure test where your AI program may be exposed across the R&D and commercial value chain, and identify practical next steps.

Book time here

Take the AI Value Readiness Assessment

The assessment gives life sciences leaders a fast, structured read on where their AI program stands against the three design imperatives that determine whether value materializes: trust, accountability, and capacity reinvestment. It also surfaces exposure against the five common design failures that prevent AI programs from moving from deployment to measurable business impact — failures that carry particular weight in regulated environments where governance, validation, and audit readiness are non-negotiable. The assessment takes minutes and provides a clear view of the gaps that matter most.

Take the assessment here

The reckoning is underway. The life sciences organizations that engage with it now will define the next era of R&D productivity and commercial performance in their sectors. The ones that wait will spend the next two years explaining where the value went.

Authors

Cecilia Edwards

Partner, AI Strategy & Transformation

Cecilia.Edwards@wavestone.com

Michael Major

Senior Manager, AI Strategy & Transformation

Michael.Major@wavestone.com

About Wavestone

Wavestone North America is a trusted partner to Fortune 1000 leaders, delivering strategic transformation and measurable business impact across North America and beyond. Backed by a global network of experts and decades of cross-industry experience, we partner with teams across the organization to modernize operations, harness the power of data and AI, strengthen cybersecurity, and drive sustainable growth.

With operations across the US and Canada, our consultants bring a disciplined, execution-focused approach that turns strategic intent into concrete results. Boasting a proven record across complex enterprise environments, we empower organizations to achieve clarity, speed, and confidence, delivering high-impact solutions that turn ambition into results amid complexity and change.

www.wavestone.com