AIDDD SUMMIT

# INSIGHTS
# THE EXECUTION GAP

# WHY AI IS TRANSFORMING DRUG DISCOVERY BUT STRUGGLING TO FIX CLINICAL TRIALS

Pharma's AI investment story has a structural flaw. Billions are flowing into drug discovery, and the results are visible: faster target identification, generative molecular design, AI-native biotech companies that did not exist a decade ago. Yet the assets those systems produce still move through a development pathway that looks, in its operational essentials, much as it did in the 1990s. The discovery engine has been rebuilt. The execution machinery has not.

The consequences are predictable with 90 percent of drug candidates that enter clinical development still failing to reach patients.

A sophisticated AI-enabled pipeline at the front end feeds into what one expert has called the "Ferrari engine with bicycle wheels" problem: discovery capabilities that are years ahead of the operational systems responsible for getting those assets through trials and into the hands of physicians.

Addressing that gap is one of the defining challenges for the pharma and biotech sector over the next decade. This report draws on expert perspectives from leading practitioners across AstraZeneca, Gilead Sciences, GSK, Pfizer, Organon, Faculty, Weave Bio, GATC Health, Advarra and Tufts University to examine where AI is genuinely moving the needle in clinical development, where the barriers remain most stubborn, and what it will take to close the gap between the technology's promise and the industry's operational reality.

**FREDDIE METHERELL**

**PORTFOLIO DIRECTOR**

# THE PROBLEM IS NOT THE TECHNOLOGY

The failure rate for AI proofs of concept in pharma is uncomfortable reading. Research from MIT published in 2025 found that 95 percent of generative AI projects are failing across industry, a worse record than the 70 percent failure rate seen for traditional machine learning initiatives over the prior decade. The parallel with drug development itself is pointed: both processes carry roughly the same odds of failure, yet companies are becoming more risk-averse toward AI investment, not less.

The reason, as Steph Skeet, Global Solutions Director at Faculty, has argued, is that the industry has consistently misdiagnosed why its AI initiatives fail. The technology is rarely the problem. Three failure modes account for most of the wreckage.
The first is starting with the technology rather than the business problem.

Organizations commission AI projects, build tools and run pilots without being able to articulate whom the technology is for, which decision it is intended to improve, or which KPI it is designed to move. The result is a proliferation of proofs of concept that allow organizations to check the innovation box while remaining disconnected from business value. As Skeet puts it: AI is everywhere except the bottom line.

The second failure mode is deploying point solutions that reinforce silos. The operational reality of clinical development involves fragmentation across three axes: time (from planning through execution across trial phases), function (clinical development, clinical operations, regulatory, supply, safety) and hierarchy (from portfolio level down to program, trial, country, site and individual patient). Point solutions deployed into discrete parts of this system deepen that fragmentation rather than resolving it. The downstream effects are concrete and costly: study designs optimized without understanding their impact on patient recruitment lead to protocol amendments; amendments lead to delays; delays compound across a portfolio in ways that no single team can see.

The third failure mode is the most underestimated: implementing technology without rewiring the process. For AI to deliver value in production rather than in pilot conditions, roles and responsibilities must change, handoffs between teams must be redesigned, decision rights must shift, and data flows must be restructured. Organizations that treat this as an afterthought find that their institutional antibodies rise up and reject the new technology. The initiative fails, the organization feels fatigued, and the case for the next AI investment becomes harder to make.


## RECOMMENDATIONS

| AVOID | SUGGESTED |
|---|---|
| • Deploying AI tools in response to board pressure without a defined business KPI<br>• Building point solutions for individual functions in isolation<br>• Treating change management as an afterthought to technology deployment | • Define the specific decision the AI is intended to support before selecting or building any technology<br>• Map fragmentation across time, function and hierarchy before designing solutions, to ensure trade-offs can be made across the system<br>• Redesign roles, handoffs, decision rights and data flows as part of the implementation, not after it |

## FROM BUSINESS INTELLIGENCE TO DECISION INTELLIGENCE

If the failure modes above share a common root, it is the tendency to think about AI as a technology to be deployed rather than a capability to be built in service of specific decisions. The discipline that addresses this directly is decision intelligence: an approach that starts with the business outcome and works backward to identify what AI is needed to support the decisions that drive it.

Tom Oliver, Head of Product at Faculty, frames the shift in straightforward terms. Business intelligence tells you what has happened. Decision intelligence tells you what will happen, why, and what you can do about it. The objective is to close decision loops faster and with greater confidence, across the full range of decisions that determine a trial's outcome.

Faculty structures this using the OUDA loop, a framework originally developed by military strategist John Boyd, who demonstrated that rapid decision iteration cycles could outperform opponents even when traditional measures of strength favored the other side. In the clinical context, OUDA stands for Observe, Understand, Decide and Act. The logic is the same: the organization that cycles through this loop faster and with better information wins.

The practical implication is a discipline of working backwards. An organization identifies the KPI it is trying to move, isolates the decision process that most influences that KPI, breaks that process into its component parts, and determines what data and AI capability are needed to support each part. This approach provides, in Skeet's description, keyhole surgery precision toward data and AI investment: targeted, sequenced and always anchored to a specific business outcome rather than a technology roadmap.

The shift also reframes what success looks like. The relevant metric is not the number of AI models built or the number of pilots launched. It is the number of high-priority business decisions that have been improved, and the degree to which better decisions in one part of the system are accelerating outcomes across the whole.

## COMPUTATIONAL TWINS: THE SIMULATION LAYER CLINICAL TRIALS NEED

One of the most consequential technical developments in clinical operations is the emergence of computational twins as a practical tool for decision-making, moving from research concept toward production deployment at a small but growing number of pharma organizations.

A computational twin is a virtual representation of an operational decision process. It is worth being precise about the distinction from digital twins, which in life sciences typically model biological systems such as organs or disease mechanisms. The computational twin models operations: how work flows, how decisions interact, how interventions in one part of a trial propagate through the system as a whole.

In practice, a computational twin brings together historical data, predictive models, optimization algorithms and business rules into a dynamic simulation. Planners can replay past trials, project forward under current assumptions, test the impact of changing inclusion/exclusion criteria, model what happens if a country is added or a site underperforms, and stress-test the portfolio against different competitive scenarios, all before committing to a course of action in the real world.

The value breaks into four measurable parameters. Speed: scenario generation that previously required ten days can be completed in ten minutes. Accuracy: in one deployment, a 21 percent uplift in forecasting performance was demonstrated. Throughput: teams can now explore hundreds of scenarios rather than tens, unconstrained by the cognitive bandwidth required to interpret each one manually. Granularity: projections that previously operated at trial level can now be disaggregated to country, site and, in developing implementations, individual patient level.

Three properties of computational twins matter particularly for clinical decision-making. Probabilistic projection surfaces the full range of possible outcomes, not just a central forecast, allowing teams to plan for best and worst cases and to build genuine trust in the outputs because the uncertainty is made visible. Cross-level coherence means all metrics derive from the same system state, so cross-functional teams can make trade-off decisions from a consistent picture rather than from competing models that contradict each other. Modularity allows different teams to work on different components while contributing toward the same objective, and in regulated environments, each decision can be traced back to its assumptions and logic for audit purposes.

The addition of AI agents extends the capability further. Conventional generative AI can only derive insight from the past. When agents are given access to a computational twin as a tool, they can run forward simulations, analyze scenarios in natural language and surface recommendations based on what will happen, not what has happened. The combination of simulation and agentic reasoning is where several practitioners see the most significant near-term operational leverage.

## AGENTS IN PRODUCTION: GILEAD'S ENROLLMENT SIMULATION SYSTEM

The most concrete operational case study available from the sector comes from Yi Hong, Associate Director of Clinical Data Science at Gilead Sciences, whose team has moved an agentic enrollment simulation system from concept to production across the company's clinical trial portfolio.

The foundation is a Poisson-Gamma statistical model, a method with a decade of validation behind it and widespread adoption across the industry. The model is Bayesian in nature: it builds priors from historical data at the start of a study and updates them continuously as actual enrollment data comes in, enabling real-time what-if analysis as the trial progresses. The statistical backbone is not new. What Gilead has built on top of it is.

Before designing the agentic layer, Hong's team mapped the existing enrollment projection process and found three structural problems. The first was accessibility: generating an enrollment projection required data scientists, biostatistics, design innovation and study team partners to coordinate across different systems and data sources, each with different levels of familiarity with the model's inputs and outputs. For two Phase 3 studies, this iteration process took approximately six months. The second was knowledge retention: insights developed within one study team were not transferable to other teams working on similar indications, meaning the same analytical work was being repeated from scratch. The third was data opacity: the model draws on both internal clinical data and purchased external data sets, and the process of generating model priors from these sources was opaque to the business stakeholders who ultimately needed to act on the outputs.

The solution is a multi-agent system built on Gilead's existing clinical data lake and AI tooling. A scoping agent receives the initial study team input, identifies knowledge gaps and produces an execution plan. A deep research agent runs multiple sub-agents including text-to-SQL agents that query internal databases in parallel, synthesizing outputs into the priors the model requires. A statistician agent calls the Poisson-Gamma optimization and forecasting tools to produce the final enrollment curve with confidence intervals. At each stage, a human-in-the-loop review allows the study team to inspect the generated insights, fill any data gaps and approve or modify assumptions before the model proceeds.

The result is a system that gives study teams direct access to data and model outputs without requiring them to understand the underlying statistical machinery. Knowledge accumulated across studies is retained and applied to future scenarios rather than being siloed within individual teams. And what was previously a six-month multi-stakeholder process is now a conversational interface.

Hong's framing of the architecture is worth noting. The agent is not the innovation. The innovations are the data lake, the validated statistical model, and the mapped workflow. The agent connects and orchestrates what already exists. Its job is to act as a value multiplier for tools and expertise that are already present, not to replace them.

## CASE STUDY: GILEAD'S ENROLLMENT SIMULATION AGENT

**THE PROBLEM**

Phase 3 enrollment projections requiring up to six months of cross-functional iteration, with knowledge siloed within individual study teams and limited business stakeholder access to underlying data and model logic.

**THE ARCHITECTURE**

A multi-agent system comprising a scoping agent, a deep research agent running parallel database queries, a statistician agent calling the Poisson-Gamma model, and human-in-the-loop review at each stage.

**THE RESULT**

Study teams have direct access to the model and its inputs. Knowledge is accumulated across studies. Enrollment scenario planning is available as a conversational interface rather than a months-long cross-functional process.

**THE PRINCIPLE**

Build the agent on existing tools, data and workflows. Connect and orchestrate what is already there. The agent's value is in removing the friction between expertise and decision, not in replacing the expertise itself.

# THE FINANCIAL RISK DIMENSION

Most AI investment in clinical development targets operational efficiency or scientific accuracy. A smaller but strategically significant set of applications targets the financial architecture of drug development itself.

Tyrone Lam, Chief Business Officer at GATC Health, argues that the 90 percent failure rate in clinical development is not primarily a scientific problem for the industry to solve. It is a capital formation problem. Investors face a portfolio where nine out of ten bets will fail.

The result is a structural constraint on the capital available to fund drug development, which limits the number of candidates that can be taken through trials in the first place.
GATC Health's D-RISQ platform generates predictive reports on how a drug will interact with its targets and perform in clinical trials, drawing on the same analytical capabilities that now underpin the first insurance product to underwrite clinical trial outcomes. The product was developed in partnership with Medical Commercial International, the exclusive syndicate of Lloyd's of London, which had sought to underwrite trial outcomes for four years but could not do so at scale without a reliable predictive model.

The commercial logic is direct. A biotech company obtains a D-RISQ report as part of the underwriting process. If the trial meets the criteria, Lloyd's issues a policy. The company can take that policy to an investor or lender as collateral: if the trial fails to meet its clinical endpoints, the lender is reimbursed in full. For investors facing a 90 percent failure rate, an insured clinical trial is a fundamentally different proposition.
The practical predictive capability was demonstrated in a case Lam described: a Phase 2 drug for liver cancer that had failed due to patient dropouts caused by a severe rash. D-RISQ analysis identified the source of the rash and determined it could be resolved by co-administering a simple antihistamine. The drug was rescued and reintroduced into Phase 2.

The broader point Lam makes is about where AI's return on investment will be most visible to the C-suite and boards that are now driving AI adoption decisions. The organizations that can show AI improving go/no-go decision speed, killing failing projects earlier, and reducing the volume of work that reaches the lab will build internal support for more ambitious AI programs faster than those pursuing harder-to-quantify scientific gains alone.

## AUTOMATING REGULATORY DOCUMENTATION

Among the near-term applications of AI in clinical development, regulatory documentation represents one of the clearest cases for adoption: high volume, highly structured, pattern-rich, and historically dependent on scarce subject matter expertise.

Lindsay Mateo, Chief Commercial Officer at Weave Bio, makes the case that regulatory workflows are, despite their complexity, fundamentally well suited to AI assistance. The process from source material to submission-ready content follows consistent patterns regardless of therapeutic area, organization size or regulatory geography. What has historically made it hard is not unpredictability but volume, coordination across global teams, and the time required of experts who have more valuable things to do than rebuild documents from scratch.

The Weave Bio platform takes source materials, including IND-enabling studies, early study reports and supporting data, and generates submission-ready content mapped to the eCTD format. It covers the full regulatory lifecycle from pre-IND to post-market, and a recently launched module handles health authority queries directly: a sponsor uploads an FDA question and receives a draft response drawn from the full document history in the system.

The critical question in any AI tool operating in a regulated environment is whether speed trades off against quality. Mateo presented data from a head-to-head study conducted with Takeda, comparing Weave-generated content against content produced by Takeda's experienced regulatory writing team across six quality dimensions. A hundred hours of manual work was reduced to three, a 97 percent reduction in time, with a first-draft quality score of 74 percent.

The 74 percent figure warrants interpretation. The goal is not to automate the document. It is to get most of the way there at the press of a button, so that the subject matter experts who previously spent weeks building from a blank page can direct their attention to the 26 percent that requires domain reasoning, contextual judgment and institutional knowledge.

LLMs are, as Mateo frames it, excellent at understanding instructions and complying with them. They are poor at reasoning from implicit domain knowledge. A fit-for-purpose wrapper that restricts inputs to verified source materials and applies document-specific templates controls for the gaps and noise that make general-purpose models unreliable in regulated authoring contexts.


The commercial model Weave Bio has developed with Parexel, in which Parexel teams build proprietary templates within the platform and receive an exclusivity window before those become available to the broader customer base, offers a practical template for how regulatory services expertise and AI platform capability can combine to create durable value for both parties.

## RECOMMENDATIONS

| AVOID | SUGGESTED |
|---|---|
| • Using general-purpose LLMs for regulatory document drafting without a fit-for-purpose wrapper<br>• Measuring AI document generation value purely by time reduction<br>• Treating AI-generated regulatory content as a finished output | • Use purpose-built platforms that restrict input to verified source materials and apply document-specific templates<br>• Evaluate first-draft quality scores and the proportion of expert time freed for high-judgment tasks<br>• Build human review and refinement into the workflow as a defined stage, not an afterthought |

## THE STATE OF PLAY: WHAT THE DATA SAYS

The most comprehensive industry-wide assessment of AI adoption in clinical development currently available comes from a 2024 research program conducted by the Tufts Center for the Study of Drug Development, published in January 2025.

The study surveyed 302 respondents across pharma, biotech and CRO organizations on AI and machine learning adoption across 36 clinical development activities, covering design and planning, trial execution and regulatory submission. It was complemented by in-depth interviews with 19 experts across 12 organizations. The headline finding is that, as of mid-2024, clinical AI adoption remains emphatically in its early stages: 30.3 percent of respondents were beginning implementation, 22.1 percent had reached partial implementation, and only 10.7 percent reported full implementation. Large biopharma companies are ahead of their smaller counterparts, and they are disproportionately building solutions in-house.

The technology adoption data from an earlier Tufts study provides important context. Across biopharma, 69 percent of organizations follow a four-step adoption process (initiation, evaluation, adoption decision, full implementation) that takes, end to end, approximately six years. The implication is that much of the AI investment visible in clinical development today will not reach full organizational deployment until the late 2020s at the earliest.

Against that backdrop, the impact data from organizations already using AI is more encouraging. Across 27 of the 36 activities studied, the average reported time reduction was 18 percent. Patient monitoring showed the most significant individual gains.

The barriers to faster adoption are, by this point, familiar but persistent. Data quality challenges were cited by 41.2 percent of respondents. Trust in AI-generated outputs came in at 37.1 percent. Intellectual property and legal concerns around data sharing affected 29.4 percent. Governance and privacy concerns were cited by 28 percent. Lack of validated AI systems was a barrier for 24 percent, and organizational design challenges for 21.7 percent.

The confidence data behind those barriers is striking. Only around one in five respondents said they were very confident in the accuracy and cleanliness of the data used to train their AI tools. Only around one in nine said they would be very likely to trust AI recommendations over a human's judgment. These figures are not primarily a technology problem: in biopharma, where data standards are high and model inputs are tightly governed, the challenge is more often that development teams cannot communicate the data quality assurance behind their models to the clinical operations colleagues who need to use the outputs. The result is a trust gap rooted in communication failure rather than actual data weakness.

The skills picture reinforces this. The most commonly cited required skills for advancing AI implementation are understanding how to integrate AI into daily work, critical thinking, and agility and iterative thinking. In each case, the same skills rank among the most frequently cited as lacking within current teams. The workforce adaptation required to realize AI's clinical development potential is running well behind the technology itself.

The study is published as: Florez, M., Do, H., Lamberti, M.J., AI Use in Clinical Development, Tufts Center for the Study of Drug Development Impact Report, January 2025; and Lamberti, M.J., Florez, M.I., Do, H. et al., The Adoption and Use of Artificial Intelligence and Machine Learning in Clinical Development, Therapeutic Innovation and Regulatory Science (2025).

## THE ORGANIZATIONAL TRANSFORMATION UNDERWAY

The technology conversation in clinical AI is, at this point, largely settled. The tools exist, the use cases are documented, and the early evidence is encouraging. The harder conversation is about the organizations that need to adopt them.
Sandeep Burugupalli, Head of Data Science, BioPharmaceuticals R&D at AstraZeneca, frames the organizational challenge in two dimensions. On the strategic side, AI can augment M&A decisions, clinical go/no-go calls and portfolio prioritization, but quantifying its impact is genuinely difficult: a month saved in a decision process has a value that is hard to isolate, particularly when the decision would likely have been made eventually without AI. On the operational side, individual efficiency gains are real but only become compelling when they accumulate across a larger workflow. Neither dimension produces the clean return on investment figures that boards and C-suite leaders are asking for.

His analogy for the current cultural moment is Tesla's Full Self-Driving. The industry is in the phase where users disengage the autonomous system and take back the wheel, not because the technology has failed but because they are habituated to manual control and have not yet learned to let go. As confidence accumulates, the letting go will happen. But Burugupalli's view is that data foundations, not cultural reluctance, remain the primary constraint. Her stated ten-year ambition is an AI agent-led clinical trial, a target she frames as requiring both organizational readiness and regulatory alignment.

Tyrone Lam offers a historical perspective on why the AI transition in pharma has taken longer than its advocates predicted. The standard approach has been to bolt AI onto existing organizational structures rather than rebuild the structures around AI. That approach fails. The pattern that works is to reorganize around the data flow, placing human experts at the points in an AI-shaped workflow where their judgment adds most value, rather than inserting AI into a workflow designed around human cognitive bandwidth. Lam describes three waves of this transition: the current co-pilot and assistant phase, a coming wave of AI sourcing in which large functions are delegated to AI systems rather than outsourced to contractors, and a third wave of AI-native organizations built from the ground up around how AI processes data. He reports that large pharmaceutical customers are targeting AI-native status within 18 to 36 months. Whether that timeline holds is uncertain. That it is the stated planning horizon is itself a signal.

Daisuke Goto, Director of Medical Affairs and Outcomes Research at Organon, raises a dimension that receives less attention in clinical AI discussions: the medical affairs function's role in the full development cycle. Medical affairs sits at the intersection of clinical evidence and physician practice, responsible for ensuring that what a drug demonstrates in trials translates into meaningful outcomes for the physicians who prescribe it and the patients who take it. The processes involved, mapping a molecule's potential meaning for different physician segments, assessing competitive landscape and payer requirements, selecting patient-reported outcomes measures that will resonate across multiple healthcare systems, are exactly the kind of evidence-synthesis and iterative analytical work that AI can accelerate significantly. Goto describes these as year-long processes today. He expects AI to shorten them substantially.

Mukul Virmani, Director of AI and Clinical Data Science at Gilead, and Stephanie Vakaljan, Director of Trial Analytics and Decision Support at GSK, both point to the same operational discipline: incremental building with explicit design for connection. The failure mode is trying to solve the whole problem at once. The approach that works is identifying a single decision to support, building the capability in a way that is designed from day one to connect to adjacent decisions upstream and downstream, and compounding value across the value chain as each connection is made.

Tom Oliver of Faculty distills the change management requirement into a single recommendation: establish cross-cutting executive sponsorship for AI programs, with representation from IT, data science and the business, and a mandate to deploy rather than govern. Governance structures that observe AI programs from a safe distance do not move organizations. Execution-focused teams with C-suite sponsorship and a brief to get AI into production do.

## CONCLUSION: THE GAP IS CLOSING, BUT SLOWLY

The evidence from across the clinical trials AI landscape points to a consistent conclusion: the technology is ahead of the organizations trying to use it, and the gap between them is an organizational and process problem, not a scientific one.
The Tufts data quantifies where the industry stands. Fewer than one in nine practitioners say they would trust AI recommendations over their own judgment. Only around one in five are very confident in the data quality behind their AI tools. Adoption timelines that have historically run to six years are only now beginning to accumulate the early evidence needed to support broader deployment.

But the direction is clear. The organizations making genuine progress share a common operating model: they start with a specific decision rather than a technology, they build on existing data and tools rather than starting from scratch, they treat process redesign as a first-order deliverable alongside technology deployment, and they accumulate small wins in ways that are explicitly designed to compound. Gilead's enrollment simulation system, Faculty's computational twin implementations, Weave Bio's regulatory documentation platform and GATC Health's clinical trial insurance product are all, in different ways, examples of this approach in practice.
The question for the sector is not whether AI will transform clinical development. It is whether organizations will build the foundations now to benefit from that transformation on a relevant timescale, or find themselves several years behind a smaller group of competitors who moved earlier and faster.

The conversations driving that question forward are what the AIDDD Summit exists to host.