IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Pharma AI Procurement: Security, Pilot and Approval Benchmarks

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-09-19 · pharma ai procurement benchmark · pharmaceutical ai procurement · ai vendor security review · pharma ai pilot · vendor risk assessment · gxp validation · ai governance · production approval · life sciences ai

Original article: https://intuitionlabs.ai/articles/pharma-ai-procurement-benchmark

In this article

Executive Summary

As of **September 19, 2026**, there is no defensible public benchmark for the full pharmaceutical artificial intelligence procurement cycle from intake through privacy, security, quality, legal, pilot, and production approval. The available evidence measures adjacent subjects. A 2025 Tufts survey gathered **302 responses** across 36 clinical-development activities (pubmed.ncbi.nlm.nih.gov), while a late-summer 2024 McKinsey survey covered more than **100 pharma and medtech leaders** (mckinsey.com). Neither publishes buyer-observed calendar time by procurement gate. Consequently, this report does not invent a median security review, pilot duration, approval rate, or pilot-to-production conversion.

The useful deliverable is a **pre-registered measurement protocol**. It defines seven mutually exclusive stages: intake, privacy and data processing agreement, security, Good Practice quality review, legal and commercial review, pilot, and production approval. Each observation records a start event, stop event, calendar days, active workdays where known, pause days, rework loops, evidence source, buyer or vendor perspective, and confidence. This approach follows the principle that preregistration occurs before data collection or analysis (help.osf.io) and that respondent identity checks should be disclosed (aapor.org). Results remain **exploratory** until coverage, buyer representation, and cell sizes support stronger language.

The practical critical path is not a fixed sequence. NIST says its AI Risk Management Framework functions can be performed in any order across the lifecycle (airc.nist.gov), and also warns that its actions are not a checklist or necessarily ordered steps (airc.nist.gov). Intake triage, privacy mapping, security architecture, quality classification, and contract issue spotting can therefore begin in parallel. A pilot should begin only after its data boundary and permitted use are approved, and production approval should depend on measured acceptance criteria, operating ownership, monitoring, change control, and rollback evidence.

The evidence burden rises with intended use. FDA and EMA identify **10 principles** for AI in drug development (fda.gov); ICH Q9(R1) says formality and documentation should be commensurate with risk (database.ich.org). For a non-GxP productivity assistant, security, privacy, legal, and operational proof may dominate. For a GxP

or regulatory-decision use, traceability, validation, independent test data, records, and controlled change become gating evidence. NIST is used here only as a taxonomy aid: its AI RMF is voluntary (nist.gov), not a certification or claim of compliance.

Introduction and Background

Pharma and biotech leaders asking for a **pharma AI procurement benchmark** usually need an operating answer, not another vendor checklist: how much calendar time to budget, where work is likely to wait, what evidence prevents rework, and which reviews can proceed together. Public research supports the urgency but not the requested clock. In one 2024 life-sciences survey, **32%** of respondents said their companies had taken steps to scale generative AI (mckinsey.com). Tufts found that **36.9%** of its sample was not yet using or implementing AI or machine learning across the activities studied (pubmed.ncbi.nlm.nih.gov). These figures describe maturity, not procurement duration.

That distinction matters. A projected implementation window, a product-development timeline, and an observed procurement-stage interval are different measures. Greenlight Guru says its 2025 medical-device report surveyed more than **500 professionals** (greenlight.guru), but its published timeline figures concern device development, not enterprise AI vendor approval. Similarly, an Arnold & Porter study surveyed **100 senior executives and department heads** with AI responsibilities (arnoldporter.com), yet reports adoption expectations rather than start and stop dates for review gates.

This report therefore publishes the design for an anonymized, buyer-centered dataset. It does not imply IntuitionLabs client experience or present consulting engagements as observations. The firm is an **adjacent advisor**, not an AI software option in this analysis. Its public service description emphasizes strategic guidance on AI adoption and technology roadmapping (intuitionlabs.ai); its stated program approach is to measure adoption, time recovered, quality, and support before scaling (intuitionlabs.ai). Those are disclosed perspectives, not benchmark results.

Benchmark Scope and Methodology

Research question and unit of analysis

The unit is one **completed or active buyer-side procurement** of an external AI capability by a pharmaceutical, biotechnology, medical-device, diagnostics, or contract-research organization. One vendor evaluated by three buyers produces three observations. One buyer evaluating three vendors produces three observations. A renewal is separate only when it triggers a substantively new risk or production decision.

The questionnaire is pre-registered before accepting responses. OSF describes preregistration as fixing the study plan before data collection or analysis (help.osf.io). Survey disclosure should also state whether identity or interview occurrence was verified, a practice named in AAPOR standards (aapor.org). The public protocol should include version history so that later changes cannot be mistaken for pre-specified analysis.

Eligibility and verification rules are:

- **Life-sciences role:** require employer type, function, seniority, and relationship to the procurement.

- **Buyer perspective:** classify buyer, vendor, advisor, or mixed; primary benchmark tables use buyer-reported records.
- **Observed versus estimated:** observed dates come from tickets, emails, contracts, or meeting records; recalled dates are marked estimated.
- **Procurement identity:** assign a random study identifier and prevent duplicate submissions for the same buyer-vendor-use-case combination.
- **Activity status:** record completed, active, paused, abandoned before pilot, ended after pilot, or approved for production.
- **Consent:** obtain explicit consent for de-identified analysis and separate consent for any quotation.
- **Geography and currency:** preserve geography and contract currency; convert spending only with a disclosed rate and date.
- **Collection period:** publish the first and last response dates and any recruitment waves.

Taxonomy and exclusions

Every case is classified as **generative AI**, **predictive machine learning**, or another algorithmic system. It is also classified by intended use: internal productivity, research support, clinical-development support, manufacturing or quality, commercial, patient-facing, or regulated-device function. EMA notes that AI used for individual clinical management may follow medical-device or in-vitro-diagnostic rules (ema.europa.eu). That path must not be pooled with a low-risk writing assistant.

GxP status is recorded as non-GxP, GxP-relevant but not a regulated record or decision, directly supporting a GxP process, or uncertain at intake. The final label must reflect quality approval, not the submitter's preference. FDA's Part 11 guidance applies to regulated electronic records across creation, modification, maintenance, archiving, retrieval, and transmission (fda.gov). That scope is one reason record impact belongs in early classification.

The benchmark excludes internal prototypes that never entered vendor intake, informal trials without buyer authorization, and vendor-estimated sales-cycle duration. Vendor responses may form a separately labeled comparison, but they cannot be merged into buyer medians. This separation is necessary because the Tufts sample itself mixed pharma and biotech companies, contract research organizations, and technology vendors (pubmed.ncbi.nlm.nih.gov).

Bias, missingness, and release thresholds

Recruitment sources, invitations, completion counts, exclusions, and missing fields must be disclosed. The CROSS survey-reporting checklist calls for explaining how missing data were handled (diposit.ub.edu). No percentage is reported without raw **n**, denominator, and missing count. No median is reported without the number of usable observations.

Publication rules are:

- **No results before responses:** blank cells remain “not yet measured,” not zero.
- **Small cells:** suppress any cross-tab cell from 1 through 10, following a conservative public-data precedent in which CMS suppresses positive values below 11 (data.cms.gov).

- **Skewed distributions:** publish median and interquartile range rather than only a mean; CDC guidance recommends median and IQR for non-normal data (cdc.gov).
- **Uncertainty:** use confidence intervals only when the design and effective sample justify them.
- **Exploratory label:** retain it when recruitment is convenience-based, buyer mix is narrow, or key segments are sparse.
- **Weighting:** do not weight results unless population targets and the weighting method are published.

Procurement Stage Model

Elapsed time must be reproducible. Each stage begins with a documented handoff and ends with a documented decision, not with an individual's impression that work “started” or was “mostly done.” Stages may overlap. The record therefore stores both duration and timestamps, allowing critical-path reconstruction rather than summing overlapping days.

Table 1 defines the seven-stage data dictionary. “Evidence captured” is the minimum needed to accept an observed date.

Stage	Start event	Stop event	Evidence captured	Primary owner
1. Intake	Complete request enters the approved intake channel	Scope, sponsor, intended use, data class, and review path are assigned	Ticket timestamps, use-case statement, sponsor, AI type, risk triage	Business sponsor and procurement
2. Privacy and DPA	Data-flow packet reaches privacy	Privacy position and data processing agreement are approved, rejected, or not required	Data map, controller or processor roles, subprocessors, retention, residency, transfer basis	Privacy and legal
3. Security	Architecture and security packet is complete	Security risk is accepted, remediated, or rejected for pilot scope	Architecture, access model, assurance reports, logging, testing, findings, exceptions	Information security
4. GxP and quality	Intended use and record or decision impact are classified	Quality approves the assurance or validation plan and release evidence	GxP classification, requirements, test evidence, traceability, change control	Quality and system owner
5. Legal and commercial	Approved term sheet and vendor paper enter review	Contract and commercial terms are executed or procurement ends	Master terms, DPA, service levels, rights, termination, audit, price, currency	Legal and procurement
6. Pilot	Authorized users receive the approved pilot configuration	Pre-specified acceptance analysis is signed	Baseline, cohort, test set, outcomes, support load, deviations, closeout	Product owner and evaluators
7. Production approval	Complete production evidence pack is submitted	Named authority approves, conditionally approves, or rejects release	Operating owner, monitoring, rollback, support, residual risk, release record	Accountable executive and control owners

The table prevents three common measurement errors. First, vendor questionnaire time is not the whole security stage if the buyer had not supplied architecture or data classification. Second, a signed contract is not production approval. Third, a pilot that ends without a production decision remains a completed pilot with an

unresolved disposition, not a successful deployment.

For every stage, capture **calendar days**, **active workdays if observable**, **pause days**, **number of return-to-vendor requests**, **number of internal rework loops**, and **reason for the longest wait**. NIST's Map function is intended to inform an initial go or no-go decision (airc.nist.gov); the benchmark treats that decision as an event, not as evidence that all later controls have been satisfied.

Security, Privacy, Legal, and Quality Evidence

Security and privacy gate

A **pharma AI vendor risk assessment** should begin with the actual system boundary. The packet needs the deployment model, hosting regions, data types, identities, privileges, integrations, model providers, subprocessors, retention, training use, logging, incident handling, and exit plan. NIST calls for procedures addressing risks from third-party software (airc.nist.gov) and says systems should be tested before deployment and regularly in operation (airc.nist.gov).

Minimum security evidence includes:

- **Architecture boundary:** components, regions, network paths, interfaces, and trust zones.
- **Identity and access:** authentication, authorization, administrative roles, service accounts, and deprovisioning.
- **Data handling:** collection, prompts, outputs, storage, retention, deletion, training use, and backups.
- **Assurance:** independent reports, testing scope, remediation status, and shared-responsibility mapping.
- **AI evaluation:** intended-use test set, failure taxonomy, human oversight, abuse tests, and acceptance thresholds.
- **Operations:** monitoring, log access, support, release notices, rollback, and termination assistance.
- **Supply chain:** model provider, hosting provider, open-source components, and material subprocessors.

The UK Information Commissioner's Office expects due diligence before procuring AI systems, datasets, or code (ico.org.uk) and recommends requesting model-development and training-process documentation (ico.org.uk). These are practical intake requirements, not elapsed-time estimates.

Privacy review should establish whether each party is a controller or processor, the permitted purpose, data categories, transfers, retention, deletion, and subprocessor approval. European Commission guidance says a processor needs a contract or other legal act (commission.europa.eu) and cannot appoint another processor without prior written authorization (commission.europa.eu). The ICO likewise recommends contracts that identify party roles and responsibilities (ico.org.uk).

Where electronic protected health information is in scope, HHS requires an accurate and thorough risk and vulnerability assessment (hhs.gov) and a written business-associate arrangement when applicable (hhs.gov). HHS also notes that customers can use a business associate agreement or service-level agreement to require additional cloud-provider evidence (hhs.gov).

Legal and commercial gate

Legal review should begin when the data flow and intended use are stable enough to draft meaningful terms, not after the technical pilot. Required positions include use rights, confidentiality, intellectual property, warranties, liability, audit, subprocessors, data return or deletion, model or service changes, support, termination, and transition. The ICO specifically recommends an exit term requiring processors to return or delete personal information (ico.org.uk) and clauses permitting audits or checks (ico.org.uk).

Security obligations belong in executable contract exhibits. CISA recommends that cloud and software-as-a-service providers make at least **six months** of security logs available without an extra charge (cisa.gov). Whether a buyer adopts that exact period is risk-specific, but log availability, format, retention, and cost should be decided before production.

GxP and quality gate

The quality path starts with intended use, not with a generic statement that a vendor is “validated.” FDA defines computer software assurance as a risk-based way to establish and maintain confidence that production or quality-management software is fit for intended use (fda.gov). ICH Q9(R1) defines quality risk management as a systematic process of assessment, control, and communication (database.ich.org).

For GxP-relevant use, the evidence pack should add:

- **Intended use and context:** users, decisions, records, environment, and prohibited use.
- **Requirements and traceability:** risk-linked requirements connected to objective tests.
- **Data provenance:** source, transformations, lineage, representativeness, and access.
- **Model evidence:** version, training boundary, independent evaluation, limitations, and acceptance criteria.
- **Record controls:** audit trail, retention, retrieval, electronic signatures where applicable, and export.
- **Change control:** configuration, model, dependency, prompt, data, and infrastructure changes.
- **Release and review:** named approvers, residual risks, periodic review, monitoring, and retirement.

EU GMP Annex 11 says a GxP computerized application should be validated and its infrastructure qualified (health.ec.europa.eu). It also calls for formal agreements when third parties provide related services or data processing (health.ec.europa.eu). EMA expects AI data sourcing, acquisition, and processing to be traceable and aligned with GxP requirements (ema.europa.eu).

AI-specific change risk can extend the approval path. EMA says an updated model may need a completely new independent test dataset after an unsatisfactory test (ema.europa.eu) and that material software, hardware, or dependency changes in high-impact uses require renewed performance evaluation (ema.europa.eu). These are reasons to record rework loops separately from first-pass review time.

Pilot, Production Approval, and Rework

A pilot is an evidence-generating stage with a bounded population, configuration, data set, duration, and decision rule. It is not free-form user access. The protocol records the baseline, intended outcome, quality and risk metrics, evaluator mix, support effort, deviations, and conclusion. NIST says AI systems should be tested

before deployment and during operation (airc.nist.gov), which makes ongoing controls part of the production decision rather than a postscript.

Pilot outcomes must use mutually exclusive labels:

- **Approved:** production scope approved without unresolved conditions.
- **Conditionally approved:** release allowed only after named conditions are met.
- **Extended:** more evidence is required under the same intended use.
- **Redesigned:** intended use, workflow, model, or deployment boundary materially changes.
- **Stopped:** buyer decides not to proceed.
- **Pending:** pilot ended, but no recorded production decision exists.

The term **engagement** is never treated as conversion. Pilot activity, weekly users, or favorable feedback may be intermediate evidence, but the numerator for pilot-to-production is only cases with a documented production approval. The denominator must identify whether active and pending cases are excluded or treated through time-to-event analysis.

Rework is coded by trigger: incomplete intake, changed use, changed data flow, missing security evidence, contract position, quality classification, failed acceptance criterion, integration constraint, operating-owner gap, or budget decision. FDA reported experience with more than **500 drug and biological-product submissions containing AI components since 2016** (fda.gov). That figure demonstrates regulatory exposure to AI, but it does not estimate procurement time or support a generic quality burden for every use case.

Production approval requires evidence beyond pilot performance:

- **Accountability:** named business, technical, security, privacy, quality, and support owners.
- **Operating controls:** monitoring, thresholds, escalation, access review, and user training.
- **Continuity:** rollback, export, vendor exit, outage handling, and manual fallback.
- **Change policy:** events that trigger notification, testing, review, or renewed approval.
- **Benefit evidence:** measured outcome, denominator, baseline, uncertainty, and support cost.
- **Release record:** exact version, scope, users, data, conditions, and approving authority.

Analysis of Key Segments

Company size and operating model

Size is captured through employee and revenue bands, but results should not assume size causes delay. Larger organizations may have specialized reviewers and standardized contracts, while smaller companies may have fewer handoffs but less dedicated capacity. Report stage distributions by band only when cells clear suppression thresholds. The organization-wide context is relevant: an IAPP survey of more than **670 people in 45 countries and territories** found **77%** of surveyed organizations were working on AI governance (iapp.org (iapp.org).

That same report placed **50%** of AI governance professionals in ethics, compliance, privacy, or legal teams (iapp.org). The benchmark should therefore record both formal owner and actual contributors, then test whether early cross-functional participation is associated with fewer loops without claiming causality.

Use-case risk and GxP status

The primary comparison is non-GxP versus directly GxP-supporting use, with “uncertain at intake” preserved as a diagnostic category. A second comparison separates internal productivity, evidence generation, operational decision support, and patient-level use. EMA's high-impact clinical-trial discussion includes model-development logs, testing logs, training data, and processing evidence in the assessable dossier (ema.europa.eu). This supports a different evidence path from an internal assistant that does not affect regulated records or decisions.

As of September 2026, the European Commission says high-risk AI embedded in regulated products has a transition period through **August 2, 2028** (digital-strategy.ec.europa.eu). The field should capture EU role and system classification, but the study must not convert an evolving legal timetable into a blanket claim that every pharma AI system is high-risk.

Deployment model

Record multi-tenant software as a service, dedicated tenant, customer cloud, on-premises, managed private deployment, and hybrid. Also record where inference, logging, support access, and backups occur. Deployment model is an explanatory variable, not a risk ranking by itself. The relevant question is whether the architecture and contract support the intended controls.

The **enterprise AI security review in pharma** should therefore compare evidence availability by deployment model: customer-controlled keys, private connectivity, administrative access, audit export, regional processing, deletion, update cadence, and rollback. No public source located in this research supplies a pharma-specific median for that review, so these fields are candidates for measurement, not implied predictors.

Data Analysis and Evidence

The public quantitative literature provides context and a model for disclosure, not the requested procurement result. Tufts reported that, averaged across 36 activities, only **10.7%** of respondents had fully implemented AI or machine learning (pubmed.ncbi.nlm.nih.gov). McKinsey found scaling steps at **32%** in its more-than-100-leader sample (mckinsey.com). These denominators, constructs, and respondent mixes differ, so the percentages should not be combined.

Budget data also measure intent, not cycle time. Among **82 respondents** in one McKinsey analysis, the share expecting annual generative-AI spending of at least **$5 million** rose from **20% for 2024 to 32% for 2025** (mckinsey.com). Procurement teams should not turn that spending distribution into a default project budget.

Deloitte's 2025 life-sciences and health-care cybersecurity survey included **323 senior cyber leaders** (deloitte.com). It reported third-party monitoring capability gaps led by lack of expertise at **56%** and insufficient tools at **54%** (deloitte.com). That finding supports collecting reviewer capacity and tooling as contextual variables, but it does not establish that either causes longer AI vendor approval.

Table 2 is the publication scorecard. All result cells intentionally remain unpopulated until verified responses exist.

Metric	Formula and denominator	Current result	Release condition
Stage duration	Median and IQR of stop date minus start date, among usable observed cases	Not yet measured	Raw n, missing n, observed or estimated split
Completion funnel	Cases reaching each gate divided by eligible intake cases	Not yet measured	Raw numerator and denominator; active cases identified
First-pass approval	Cases with zero rework loops divided by decided cases at that gate	Not yet measured	Decision and loop coding complete
Pilot outcome	Each mutually exclusive outcome divided by pilots with a recorded disposition	Not yet measured	Pending and active cases reported separately
Pilot-to-production	Production approvals divided by eligible completed pilots	Not yet measured	No engagement proxy; time window disclosed
Evidence frequency	Cases requesting each artifact divided by cases answering the item	Not yet measured	Missing responses shown; small cells suppressed
Segment difference	Distribution or proportion by GxP, size, use risk, and deployment	Not yet measured	Each displayed cell has n of at least 11
Rework matrix	Loop count by stage and coded trigger	Not yet measured	Multiple triggers and unresolved cases disclosed

This blank scorecard is a feature, not a missing calculation. Publishing zeros would falsely imply observed absence. Publishing a median from recalled sales-cycle anecdotes would mix units and perspectives. Once data exist, non-normal durations should be summarized with median and interquartile range, consistent with CDC guidance (cdc.gov).

The benchmark analogue also shows why method detail matters. Editorial coverage of a Greenlight Guru survey reproduced the respondent mix (mpo-mag.com) and fielding window (mpo-mag.com). Those details make a figure auditable. The proposed procurement study should disclose the same class of information directly from its own study records, not depend on later editorial reconstruction.

Critical-Path Planning and Parallel Work

The planning worksheet uses user-entered durations. It has no default benchmark. For each stage, enter start date, earliest possible start, dependency, owner, elapsed days, pause days, and confidence. The critical path is the longest chain of dependent work, not the sum of all seven stage durations.

Table 3 distinguishes work that can normally start together from decisions that require an upstream artifact. Local policy and use-case risk still govern each case.

Workstream	Can begin when	Can run in parallel with	Hard dependency before close
Intake triage	Sponsor and intended use are named	Initial market scan and budget framing	Data class, deployment boundary, GxP hypothesis

Workstream	Can begin when	Can run in parallel with	Hard dependency before close
Privacy and DPA	Preliminary data flow and parties are known	Security architecture and contract issue list	Final subprocessors, regions, retention, party roles
Security	Architecture and pilot boundary are supplied	Privacy, quality classification, commercial negotiation	Resolved findings, operating controls, accepted residual risk
GxP and quality	Intended use and record or decision impact are described	Security and pilot protocol design	Approved requirements, evidence, traceability, change plan
Legal and commercial	Data and service model are stable enough to draft	Technical reviews and pilot planning	Approved DPA, security schedule, service and exit terms
Pilot setup	Pilot data, users, metrics, and safeguards are approved	Contract completion where policy permits	Authorized environment and signed acceptance protocol
Production approval	Pilot closeout and operations pack are complete	Final training and release scheduling	All mandatory control-owner decisions and release record

The central scheduling move is to front-load artifacts shared by multiple reviewers: intended use, system context, data flow, deployment architecture, subprocessor list, record impact, and pilot acceptance criteria. NIST's four functions are Govern, Map, Measure, and Manage (airc.nist.gov), but the framework permits them in different orders (airc.nist.gov). That makes it a useful taxonomy for parallel work, not a claim that following it guarantees approval.

Worksheet procedure:

- **Step 1:** name the decision and intended use in one testable sentence.
- **Step 2:** draw the data, identity, model, integration, and operating boundary.
- **Step 3:** classify GxP, regulated-record, personal-data, health-data, and device impact.
- **Step 4:** identify artifacts consumed by two or more review functions.
- **Step 5:** start independent reviews when their minimum inputs are complete.
- **Step 6:** log every pause and distinguish buyer wait from vendor wait.
- **Step 7:** record rework cause, requester, request date, response date, and closure.
- **Step 8:** calculate the realized critical path from timestamps after the decision.

Benchmark Questionnaire and Data Dictionary

The downloadable study package should contain the questionnaire, coding manual, anonymized aggregate table, change log, and analysis script. The questionnaire asks factual dates before opinions and displays “unknown” rather than forcing a guess.

Core questionnaire fields are:

- **Organization:** type, geography, employee band, revenue band, and regulated markets.
- **Respondent:** function, seniority, buyer or vendor relationship, and procurement role.
- **System:** AI type, model source, deployment model, intended use, users, and integrations.
- **Risk:** data types, GxP status, record impact, device status, and decision impact.

- **Timeline:** each stage's start, stop, status, evidence source, pause days, and confidence.
- **Evidence:** artifacts requested, supplied, rejected, waived, and accepted with conditions.
- **Rework:** loop count, trigger, initiator, affected stage, and added elapsed days.
- **Pilot:** baseline, success measures, sample, duration, outcome, support effort, and deviations.
- **Production:** decision, scope, conditions, accountable owner, monitoring, and rollback.
- **Commercial:** currency, contract band, pricing basis, and whether price includes pilot services.
- **Disclosure:** recruitment source, consent, quotation consent, and follow-up permission.

The data dictionary defines every value and prohibited inference. “Security complete” requires a recorded risk decision. “Pilot approved” is not “production approved.” “Estimated” never silently becomes observed. Buyer-reported and vendor-reported dates remain separate. Missing does not mean no. A stage with start but no stop is right-censored or active, not assigned zero days.

The final release should publish missingness by variable and respondent mix by function. It should also state whether the sample is convenience-based. A credible benchmark can be useful while exploratory, but only if readers can see exactly who answered, which stages were observed, and where the data are thin.

Implications and Future Directions

For CIOs and AI program leaders, the immediate implication is to budget for **evidence production and reviewer capacity**, not just software and pilot labor. Deloitte's cyber survey found capability gaps in third-party monitoring at 56% for expertise and 54% for tools (deloitte.com). A future procurement benchmark should test whether those capacity variables correlate with waiting time, while avoiding a causal conclusion from cross-sectional data.

For security, privacy, legal, and quality teams, the protocol creates a shared clock. Review rigor is not measured by elapsed days alone. A shorter review can reflect better intake evidence, narrower scope, more reviewer capacity, or weaker scrutiny. The dataset therefore pairs time with requested evidence, rework, decision, and risk class.

For vendors, the practical requirement is a reusable evidence pack that stays aligned with the deployed configuration. Model and training documentation are specifically identified in ICO due-diligence guidance (ico.org.uk). For high-impact medicinal-product uses, EMA expects detailed traceability (ema.europa.eu). Evidence that does not match the actual tenant, model, region, or subprocessors should not be coded as complete.

Future releases should add enough observations to compare GxP status, deployment model, organization size, geography, and use-case risk without exposing respondents. Longitudinal follow-up can test whether mature governance reduces rework. Any revision should retain original definitions or bridge old and new measures, preserving trend interpretability.

Frequently Asked Questions (FAQs)

How long does a pharma AI vendor security review take?

No authoritative public source located for this report publishes a pharma-specific median with buyer-observed start and stop events. The correct current answer is **not yet measured**. Organizations should enter their own stage times in the critical-path worksheet and distinguish elapsed time, pause time, and rework.

What is the pharmaceutical AI procurement process?

The measured process has seven stages: intake, privacy and DPA, security, GxP and quality, legal and commercial, pilot, and production approval. They overlap when inputs permit. NIST says its risk functions can occur in any order across the lifecycle (airc.nist.gov).

How long should a pharma AI pilot run?

There is no defensible universal duration. A **pharmaceutical AI pilot approval process** should evaluate a pre-specified population, operating conditions, acceptance criteria, and support burden. NIST expects testing before deployment and regularly in operation (airc.nist.gov). Duration should be recorded as observed calendar time, not borrowed from a different industry's implementation estimate.

What determines an AI vendor approval timeline in pharma?

The protocol tests, rather than assumes, the influence of intended use, data types, GxP status, deployment model, evidence completeness, reviewer capacity, contract positions, integrations, pilot results, and rework. ICH Q9(R1) supports making effort and documentation commensurate with risk (database.ich.org).

Is NIST AI RMF compliance required?

This report uses NIST AI RMF only to organize risk work. NIST describes the framework as voluntary (nist.gov). It is not presented here as certification, legal advice, or proof of regulatory compliance.

Conclusion

The honest 2026 finding is that a **life-sciences AI procurement benchmark does not yet exist in publishable public form** for the requested seven-stage cycle. Adjacent surveys quantify implementation across 36 activities (pubmed.ncbi.nlm.nih.gov), governance across 45 countries and territories (iapp.org), budgets, and cybersecurity capacity, but none supplies verified buyer-side start and stop events across intake, privacy, security, quality, legal, pilot, and production approval. Converting those studies into a timeline would create false precision.

The proposed protocol makes the missing benchmark measurable. It fixes the unit of analysis, stage boundaries, respondent verification, buyer and vendor separation, GxP and use-risk classifications, missing-data rules, small-cell suppression, and pilot outcome definitions before data arrive. It also provides a critical-path worksheet that uses organization-specific inputs and recognizes parallel review.

Decision makers can apply the framework immediately without pretending that blank evidence is a number. Start shared artifacts early, measure every handoff and pause, require a recorded disposition, and treat production approval as distinct from pilot engagement. When verified responses are sufficient and representative enough, the blank scorecard can be replaced by medians, interquartile ranges, raw denominators, uncertainty intervals where defensible, and clearly labeled segment comparisons. Until then, **“not yet measured” is the benchmark's most important result**.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://pubmed.ncbi.nlm.nih.gov/40439837/
2. https://www.mckinsey.com/industries/life-sciences/our-insights/scaling-gen-ai-in-the-life-sciences-industry
3. https://help.osf.io/article/330-welcome-to-registrations
4. https://aapor.org/standards-and-ethics/disclosure-standards/
5. https://airc.nist.gov/airmf-resources/airmf/5-sec-core/
6. https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development
7. https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_Step4_2025_0115.pdf
8. https://www.nist.gov/itl/ai-risk-management-framework
9. https://intuitionlabs.ai/articles/pharma-ai-procurement-rfp-template-scorecard
10. https://www.greenlight.guru/state-of-medical-device
11. https://www.arnoldporter.com/en/perspectives/topics/the-convergence-of-life-sciences-and-artificial-intelligence
12. https://intuitionlabs.ai/
13. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
14. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application
15. https://diposit.ub.edu/bitstreams/aac5ba28-af7b-4516-bb11-3c1740be7749/download
16. https://data.cms.gov/sites/default/files/2023-09/ee7944fc-a460-4b3c-8039-6526a68139e0/ETC-MY1-MC-DataDictionary-508.pdf
17. https://www.cdc.gov/pcd/for_authors/top_five.htm
18. https://intuitionlabs.ai/articles/pharma-ai-vendor-due-diligence-checklist
19. https://ico.org.uk/for-organisations/advice-and-services/audits/data-protection-audit-framework/toolkits/artificial-intelligence/contracts-and-third-parties/
20. https://commission.europa.eu/law/law-topic/data-protection/information-business-and-organisations/application-gdpr_en
21. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html
22. https://www.hhs.gov/hipaa/for-professionals/faq/2084/do-the-hipaa-rules-require-csps-that-are-business-associates-to-provide-documentation-or-allow-auditing-of-their-security-practices-by-their-customers-who-are-covered-entities-or-business-associates/index.html
23. https://www.cisa.gov/sites/default/files/2024-08/SecureByDemandGuide_080624_508c.pdf
24. https://intuitionlabs.ai/articles/csa-vs-csv-validating-ai-systems-fda-guidance

25. https://www.fda.gov/media/188844/download
26. https://intuitionlabs.ai/articles/risk-based-ai-validation-pharma-ich-q9
27. https://intuitionlabs.ai/articles/pharma-ai-validation-evidence-fda-ema
28. https://health.ec.europa.eu/document/download/8d305550-dd22-4dad-8463-2ddb4a1345f1_en
29. https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions?trk=article-ssr-frontend-pulse_little-text-block
30. https://intuitionlabs.ai/articles/agentic-ai-pharma-scaling-production
31. https://iapp.org/resources/article/ai-governance-profession-report
32. https://intuitionlabs.ai/articles/pharma-ai-governance-gxp-compliance
33. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
34. https://www.mckinsey.com/featured-insights/charts/ai-budgets-grow-in-life-sciences
35. https://www.deloitte.com/us/en/services/consulting/articles/life-sciences-health-care-ciso-survey.html
36. https://www.mpo-mag.com/breaking-news/survey-identifies-challenges-to-improving-product-development-processes/

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai