GxP
Frame

WHITE PAPER · VERSION 4.0

Governing AI in Pharmaceutical Operations

Version 4.0 · Regulatory status last reviewed September 21, 2026

An end-to-end operating method for deciding whether AI-supported quality work may proceed: the AI-Supported Work Authorization Cycle.

By Brian Drapeau, GxP Frame September 20, 2026 Time of read: ~26 minutes

A practical operating method for pharmaceutical quality leaders deciding whether a defined AI-supported workflow may proceed, what conditions apply, and what evidence requires restriction or withdrawal.

CONTENTS

What must be true before the work proceeds....2
What industry experience tells us....3
The control basis the cycle sits on....4
The AI-Supported Work Authorization Cycle....8
Stage 0: Inventory and ownership....12
Stage 1: Define the work and risk....12
Stage 2: Set authority and preserve decision evidence....14
Stage 3: Test the complete workflow....16
Stage 4: Authorize routine use and monitor....21
Stage 5: Change, correct, reauthorize, and retire....22
Count the time and cost of the whole job....23
Putting the cycle to work in a fictional manufacturer scenario....24
How we would know the approach is not helping....25
Decide whether the work can proceed....25
Sources....26
How to cite this paper....27
Version 4.0 changes....27
AI Work Authorization Toolkit....28

EXECUTIVE SUMMARY

What work is authorized now, under what limits, on what evidence, and what finding would require the organization to restrict or stop it?

A defensible authorization decision applies to a defined workflow, not AI in the abstract. The AI-Supported Work Authorization Cycle begins by inventorying the use and assigning ownership. It then defines the work and risk, sets an authority envelope, tests the complete human-AI workflow on representative cases, authorizes only the scope supported by evidence, and reconsiders the decision when the workflow changes or fails.

The method is designed for pharmaceutical manufacturing and quality operations. It works alongside the lifecycle-oriented, risk-based practitioner guidance in GAMP Guide: Artificial Intelligence and GAMP 5, while adding a local authorization decision for each configured human-AI workflow. It does not replace existing GMP responsibilities, validation, data-integrity controls, Part 11 analysis, or quality-unit accountability.

Public examples support the need to manage evidence access, human judgment, and change, but each organization must establish the evidence and controls appropriate to its own work. GxP Frame's reference toolkit provides a use inventory, Gate Decision Record, workflow scoring sheet, AI Authorization Register, Independent Effectiveness Review Record, implementation guide, and role cards that turn the cycle into a practical starting point.

The method states acceptance criteria in terms an auditor can check. For the illustrative deviation-chronology pilot, authorization requires zero critical omissions across at least 59 independent, representative holdout cases, supporting a true critical-omission rate no higher than about 4.95% at 95% one-sided confidence. For the same 5% claim, one critical omission requires at least 93 total cases with no further failures; demonstrating a rate below 1% requires 299 zero-failure cases.

Start Monday with Stage 0: inventory actual AI use, including unapproved use; identify the systems and records involved; and assign a process owner, system owner, data owner, and quality owner before deciding what may proceed.

Readers who want the method in one record can start with the illustrative Gate Decision Record after Figure 1, then continue to the full fictional-manufacturer walkthrough.

What must be true before the work proceeds

Consider a deviation summary prepared by AI. The chronology looks sound, the documents are linked, and a reviewer approves it. Months later, someone finds an equipment record that contradicts the conclusion. The summary left it out. To understand this illustrative failure, we need to know which records the system could access, how it selected them, what the reviewer could see, and why the summary was approved.

AI must not support pharmaceutical quality work until the organization has defined the task, established what the system may do, and tested whether the complete human and AI workflow can perform representative work reliably. The resulting evidence must determine whether the work may proceed, what limits apply, and what findings require correction, restriction, or withdrawal.

This approach connects four questions that organizations often address separately: whether the evidence is adequate, whether the system has appropriate authority, whether people can perform their assigned role, and whether the organization can find and correct failures. Passing a model test or completing a course does not establish that the combined workflow will produce acceptable work. A qualified person can still miss a record that changes the conclusion, especially when the system controls which evidence reaches the reviewer.

I developed this operating method through earlier work on manufacturing readiness, governance by design, AI qualification, and agent authority. The method treats the relationship among the person, system, version, use case, source environment, and granted authority as the operating unit.

The focus here is pharmaceutical manufacturing and quality operations. Other GxP settings require controls suited to their records, decisions, and governing requirements.

What industry experience tells us

Industry experience supports the need to evaluate the work around the technology. Roche describes trusted data, employee development, process redesign, and governance as connected conditions for useful AI outcomes. McKinsey's pharmaceutical analyses similarly place talent and change management alongside strategy, technology, analytics, and data, and call for operators and technical staff who can recognize errors and make informed decisions. [1–2]

GxP
Frame

Lilly's responsible-use disclosure assigns employees accountability for AI outputs and describes evaluation for specific uses. GSK reports generative AI across manufacturing sites for examining historical investigations, alongside further inspection-readiness work. These accounts show organizations confronting evidence access, human judgment, and workflow design together. [3–4]

The useful question is whether an organization can use available evidence to make a defensible decision about a particular AI-supported job.

The control basis the cycle sits on

The cycle sits inside existing quality responsibilities, record requirements, and lifecycle controls. It does not replace them, and no source below establishes that completing these six stages is sufficient authorization.

CONTROL BASIS AND LIMITS

Quality-unit responsibility

WHAT IT REQUIRES OR SUPPORTS HERE

21 CFR 211.22(c) requires the quality control unit to approve or reject all procedures and specifications impacting drug-product identity, strength, quality, and purity. In Warning Letter 320-26-58 (April 2, 2026), FDA found that a homeopathic drug manufacturer "used AI to create drug product specifications, procedures, and master production or control records" and stated that the failure to review those documents for accuracy and CGMP compliance violated 21 CFR 211.22(c). [5]

WHAT IT DOES NOT ESTABLISH

A universal AI certification, permission to outsource the regulated decision, or any conclusion about the adequacy of a mature governance program. The cited firm had broad baseline CGMP failures, including violations of 21 CFR 211.165(b) and 211.84(d).

Qualified personnel

WHAT IT REQUIRES OR SUPPORTS HERE

21 CFR 211.25 requires education, training, experience, or their combination to enable assigned functions, with continuing training as needed. [6]

WHAT IT DOES NOT ESTABLISH

That course completion or awareness training demonstrates performance in a configured human-AI workflow.

GxP Frame

Quality risk management

WHAT IT REQUIRES OR SUPPORTS HERE

ICH Q9(R1) supports effort, formality, and documentation proportionate to risk, uncertainty, importance, and complexity across a lifecycle. [7]

WHAT IT DOES NOT ESTABLISH

That risk language can excuse an unacceptable practice or unsupported conclusion.

Pharmaceutical quality system

WHAT IT REQUIRES OR SUPPORTS HERE

ICH Q10 connects management responsibility, knowledge management, monitoring, CAPA, change management, and management review. [8]

WHAT IT DOES NOT ESTABLISH

That this particular six-stage cycle is a regulatory requirement.

Electronic equipment and records

WHAT IT REQUIRES OR SUPPORTS HERE

Where applicable, 21 CFR 211.68 supports authorized changes, checks proportionate to system complexity and reliability, and backup safeguards. [9]

WHAT IT DOES NOT ESTABLISH

That every prompt or output is a regulated record or that an AI use is acceptable.

Data integrity and provenance

WHAT IT REQUIRES OR SUPPORTS HERE

FDA data-integrity guidance and PIC/S PI 041-1 support reliable, accurate, complete, consistent, enduring, and available records, with data governance based on process and technology risk. [10–11]

WHAT IT DOES NOT ESTABLISH

That a fluent answer or a citation list proves that the evidence set was complete.

GxP
Frame

Human-AI interaction

WHAT IT REQUIRES OR SUPPORTS HERE

Joint FDA and EMA good-AI-practice principles support examining the complete system, including human interaction, provenance, multidisciplinary expertise, and continuing performance. [12]

WHAT IT DOES NOT ESTABLISH

A statutory qualification scheme or endorsement of this method.

European AI proposals — Status as of September 21, 2026: draft

WHAT IT REQUIRES OR SUPPORTS HERE

The effective EudraLex listing must be distinguished from the consultation draft for Annex 22. Consultation ran July 7 to October 7, 2025, and is closed. EMA held a multistakeholder workshop on June 30 and July 1, 2026, to gather expert input and stated that it was still considering the consultation results. No publication or application date has been announced. [13–14]

WHAT IT DOES NOT ESTABLISH

That draft Annex 22 is effective law or that adding a reviewer makes a critical function non-critical.

GAMP lifecycle guidance

WHAT IT REQUIRES OR SUPPORTS HERE

GAMP Guide: Artificial Intelligence and GAMP 5 provide lifecycle-oriented, risk-based practitioner guidance for GxP systems. [15–16]

WHAT IT DOES NOT ESTABLISH

That GAMP requires these six stages or these specific records.

AI governance systems

WHAT IT REQUIRES OR SUPPORTS HERE

NIST AI RMF 1.0 calls for AI inventories, defined roles, accountability, monitoring, and documented risk decisions. ISO/IEC 42001:2023 provides a management-system structure for responsible AI risk and opportunity management. [17–18]

WHAT IT DOES NOT ESTABLISH

A pharmaceutical authorization decision, GMP compliance, or acceptable performance for a specific use.

GxP
Frame

How the authorization cycle relates to GAMP

This is a conceptual crosswalk. GAMP supplies lifecycle-oriented, risk-based practitioner guidance; the cycle adds a local gate decision for a configured human-AI workflow. [15–16]

CONCEPTUAL GAMP CROSSWALK

0. Inventory and ownership

RELATED LIFECYCLE CONCERN
Governance, system inventory, intended use, supplier and organizational responsibilities

ADDED LOCAL DECISION
Register, contain, prohibit, or assign named ownership for each actual use

1. Define work and risk

RELATED LIFECYCLE CONCERN
Requirements, intended use, process understanding, risk assessment, and lifecycle planning

ADDED LOCAL DECISION
Approve, narrow, or reject the specific human-AI workflow scope

2. Set authority and records

RELATED LIFECYCLE CONCERN
Security, access, data integrity, electronic records, interfaces, and control design

ADDED LOCAL DECISION
Authorize bounded read, recommend, change, or approval functions, or prohibit them

3. Test the workflow

RELATED LIFECYCLE CONCERN
Risk-based verification, test evidence, acceptance criteria, traceability, and supplier evidence

ADDED LOCAL DECISION
Authorize a bounded pilot, correct, narrow, or stop based on representative work

4. Authorize and monitor

RELATED LIFECYCLE CONCERN
Release, operation, performance monitoring, incident signals, and periodic review

ADDED LOCAL DECISION
Authorize routine use, continue, restrict, investigate, suspend, or expand

5. Change, correct, retire

RELATED LIFECYCLE CONCERN
Change control, problem management, corrective action, re-verification, continuity, and retirement

ADDED LOCAL DECISION
Reinstate, narrow, requalify, stop, or retire the workflow

The AI-Supported Work Authorization Cycle

The following cycle turns the argument into a set of gate decisions. Each stage produces evidence for proceeding, narrowing the use, correcting it, or stopping it.

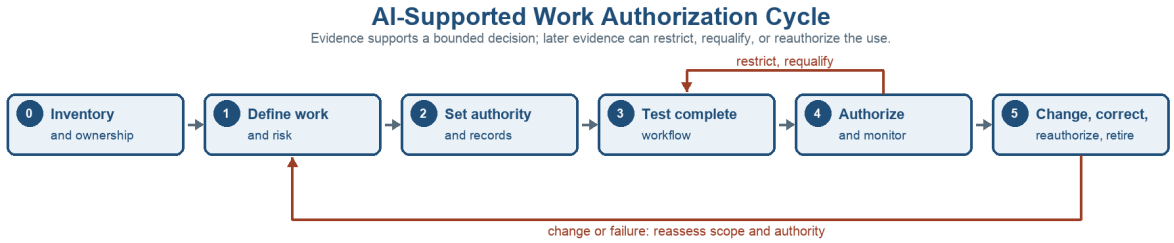


Figure 1. The AI-Supported Work Authorization Cycle.

What "tested" has to mean

Zero critical omissions in 59 independent holdout cases supports a true rate no higher than about 4.95% at 95% one-sided confidence. Not "we ran some examples."

Each gate must have a named decision maker and a recorded outcome. Pilot authorization does not authorize routine use, and routine-use authorization does not authorize expansion to a new task, source, site, or authority level. The Gate Decision Record must identify the evidence reviewed, accepted conditions, unresolved risks, accountable decision maker, and trigger for reconsideration. Completed activities are not authorization.

GxP
Frame

This record is one of seven artifacts in the AI Work Authorization Toolkit, alongside the use inventory, workflow scoring sheet, AI Authorization Register, Independent Effectiveness Review Record, implementation guide, and role cards. See the toolkit: https://www.gxpframe.com/ai-work-authorization-toolkit/

Illustrative Gate Decision Record: Deviation Chronology Pilot at a Fictional Manufacturer

This filled example is fictional. The controls and decision illustrate the method; they do not describe any organization's internal implementation or results.

Decision

Authorize a bounded pilot.

System/version/configuration

Chronology Assist v2.7.4; retrieval configuration DC-04; approved read-only deviation repository.

Date

September 20, 2026.

Review by

December 20, 2026, or earlier if a stop or restrict trigger occurs.

Authorized scope

Prepare a cited chronology from an approved, read-only collection of deviation records and associated investigation responses for an investigator who verifies every consequential statement.

Evidence reviewed

Zero critical omissions across 59 independent, representative holdout cases; source-link verification; reviewer decisions; false-connection results; review time; negative permission tests; and recovery tests.

Open risks

Unfamiliar terminology, translation history, incomplete indexing, corpus changes, and review workload may still conceal a consequential record.

Conditions

Read-only access; cited source passages; visible search limits; no automated conclusion, approval, or record change; quality-owner review; retained query, source, configuration, output, and decision evidence.

Stop or restrict trigger

Any critical omission, unavailable required source, unsupported cross-site conclusion, material corpus or configuration change, or review queue that prevents meaningful verification.

Decision maker

Quality owner, with documented input from the process owner, system owner, data owner, and evaluation owner.

GxP
Frame

THE AI-SUPPORTED WORK AUTHORIZATION CYCLE

0. Inventory and ownership

QUESTION
What AI is in use, including unapproved use, and who owns each use?

REQUIRED EVIDENCE
Use inventory, systems and data accessed, users, suppliers, named process, system, and quality owners

ACCOUNTABLE OWNER
AI governance lead and quality unit

GATE DECISION
Register, contain, prohibit, or assign ownership

1. Define the work and risk

QUESTION
What task and decision will AI support, and what could failure affect?

REQUIRED EVIDENCE
Intended use, source boundary, applicable requirements, baseline process, risk assessment

ACCOUNTABLE OWNER
Process owner and quality owner

GATE DECISION
Approve the scope, narrow it, or reject it

2. Set authority and records

QUESTION
What may the system read, recommend, change, or approve, and which records must be retained?

REQUIRED EVIDENCE
Permissions, prohibited actions, approval design, record and signature requirements, retention plan

ACCOUNTABLE OWNER
System owner, quality owner, IT, and data owner

GATE DECISION
Authorize bounded functions or prohibit them

3. Test the complete workflow

QUESTION
Can the system and people perform representative work under realistic conditions?

REQUIRED EVIDENCE
Challenge cases, system results, reviewer performance, omissions, false alarms, time, and recovery tests

ACCOUNTABLE OWNER
Evaluation owner and quality unit

GATE DECISION
Authorize a pilot, correct, narrow, or stop

4. Authorize routine use and monitor

QUESTION
Does pilot evidence justify routine use, and does performance remain acceptable?

REQUIRED EVIDENCE
Pilot results, unresolved risks, incidents, near misses, overrides, workload, outcome measures, drift indicators, and user feedback

ACCOUNTABLE OWNER
Process owner and quality unit

GATE DECISION
Authorize routine use or expansion, continue, restrict, investigate, or suspend

5. Change, correct, and retire

QUESTION
What changed, what was affected, and did the response work?

REQUIRED EVIDENCE
Change impact, regression results, requalification, corrective-action evidence, records, dependencies, and continuity plan

ACCOUNTABLE OWNER
Change owner, system owner, and quality unit

GATE DECISION
Reinstate, narrow, stop, or retire

Stage 0: Inventory and ownership

Stage 0 matters because an organization cannot govern a use it has not identified. The inventory must describe actual work, not merely list vendors. "AI used in quality" is too broad. Each entry must identify the task, source systems, output, users, downstream decision, and whether the system can act. Unapproved consumer-tool use belongs in the same inventory process because it may expose regulated information or create outputs that enter the quality system without an accountable owner.

Ownership must be divided clearly. The process owner defines the work and acceptable outcome. The system owner maintains the configured service, integrations, access, and technical evidence. The data owner governs the sources and permitted uses of information. The quality unit determines whether the proposed controls and evidence support the regulated use and retains its existing approval responsibilities. An AI governance body can set common policy, coordinate expertise, and resolve cross-functional questions, but it must not become a substitute owner for every local process. Suppliers provide evidence and change information; they do not accept the manufacturer's regulated decision on its behalf.

The inventory must remain connected to procurement, access management, incident reporting, and retirement. Otherwise, it becomes a periodic spreadsheet that misses the work it was created to govern. New integrations, expanded permissions, unexpected outputs in controlled records, and reports of consumer-tool use must trigger inventory review rather than wait for an annual reconciliation.

The cycle does not require the same ceremony for every use. A bounded formatting tool and an agent that can alter a regulated record present different questions. Evidence, documentation, and review should follow the significance, uncertainty, and complexity of the work.

Stage 1: Define the work and risk

I use the shorthand person × system × version × use case. For an agent, add the source environment and authority envelope: the systems and records it may access, the actions it may take, the approvals it needs, and the conditions that stop it. Readiness depends on this complete relationship and can change when any part changes. [19]

"Prepare a draft chronology from a defined set of approved deviation records for an investigator who verifies each consequential statement" can be evaluated. "AI for Quality" cannot. A useful intended use identifies:

- **Task.** The bounded work the system supports.
- **Source boundary.** The systems, records, and data it may use.
- **Expected output.** The artifact or recommendation it may produce.
- **User.** The role responsible for operating or reviewing it.
- **Downstream decision.** The decision the output may inform.
- **Prohibited actions.** What the system may not do.

The intended use must also identify how failure could affect product quality, patient safety, data integrity, compliance, or availability.

The system description must include the interface, connected tools, retrieval sources, configurations, permissions, and points where a person must intervene. A capable reviewer may face a materially different job after a retrieval change, model update, new product, altered interface, or grant of write access. The name of the product may remain unchanged while the qualified state has changed.

Define the baseline process at the same time. If the current human process already misses records, produces inconsistent classifications, or takes an unacceptable amount of time, the comparison should expose that weakness. The objective is not to preserve a flawed baseline. It is to determine whether the proposed workflow produces acceptable work and whether its benefit justifies the new risks and controls. A comparison that measures only the AI condition can misattribute pre-existing failures to the tool or conceal failures that both methods share.

Different technologies require different evidence. A classifier used in visual inspection, a predictive model used in process analysis, and a language model drafting an investigation create different failure modes. For a classifier, evaluation should address independent test data, important defect classes, subgroup performance, calibration, and the consequences of false positives and false negatives. A strong average can conceal poor performance on a rare but consequential defect.

For generative systems, verification must address meaning. A fluent sentence can reverse a negation, remove a qualification, or imply causation where the source shows only sequence. Agreement between repeated answers gives limited reassurance when the same missing source or unsupported inference drives each answer. Tested calculation tools must perform controlled calculations, deterministic rules must enforce permissions, retrieval must identify

candidate evidence, and generative AI may organize or draft material within those boundaries. Test each component for its assigned job and then test the complete workflow.

The product and task also shape the challenge cases. In QC, a reviewer may need to find omitted raw data, a dilution error, a method-version mismatch, or an unsupported explanation for an out-of-specification result. In technology transfer, the risk may be extrapolation beyond studied conditions. In patient-specific manufacturing, identity and custody require explicit testing because a plausible summary attached to the wrong patient or material is a serious failure. For external partners, the system must preserve site identity, supplier status, quality-agreement responsibility, and provenance across organizational boundaries.

Stage 2: Set authority and preserve decision evidence

As AI moves from producing content to taking action, the organization needs a precise account of what each system may observe, interpret, recommend, decide, execute, and approve. Capability describes what the technology can do. Permission describes what the organization has authorized it to do. The Authority Envelope must contain the following elements. [20]

Element	Recorded where
Accessible systems and records	Gate Decision Record Authority Envelope
Allowed operations	Gate Decision Record Authority Envelope
Prohibited actions	Gate Decision Record Authority Envelope
Approval points	Gate Decision Record Authority Envelope
Stop conditions	Gate Decision Record Authority Envelope and monitoring plan
Accountable person	Gate Decision Record decision and approval fields

Part 11 should be applied with equal precision. It does not make every AI prompt or output a Part 11 record. Under FDA's scope guidance, Part 11 applies when records required by predicate rules are maintained electronically in place of paper or relied upon for regulated activities, when designated records are submitted electronically to FDA, and when electronic signatures are intended to serve as required signatures. The organization should determine which records and signatures fall within that scope. For applicable systems, authority checks, operational checks, access controls, accountability for electronic signatures, and reliable records are directly relevant to AI-supported approvals and actions. [21]

Where 21 CFR 211.68 applies to automatic, mechanical, or electronic equipment used in drug manufacturing, subsection (b) requires appropriate controls so that changes to master

production and control records or other records are instituted only by authorized personnel. It also addresses checks on computer input and output according to system complexity and reliability, together with backup safeguards. These requirements support the authority envelope and record design, but they do not prescribe this paper's workflow or establish that an AI use is acceptable. [9]

Make that determination here, before a system has generated months of records. The team must identify which artifact is the regulated record, which supporting evidence must remain linked to it, whether an electronic approval is intended to have the force of a required signature, and what information must be available during an investigation or inspection. An AI interaction may be transient in one workflow and consequential evidence in another. The applicable predicate rule and actual business reliance determine the treatment.

These limits require technical enforcement. An instruction telling an agent not to approve a record provides weak protection if its credentials permit approval. The agent must have its own attributable identity and only the access required for its assigned function. It must not inherit all the privileges of the employee who launches it.

Approval must remain tied to the state reviewed. If someone approves an update and the underlying record changes before the agent acts, the system must determine whether the approval still applies. Where the change undermines the original basis, the action requires another review. Permission to retrieve records does not confer permission to determine root cause, open a corrective action, or approve closure.

GxP
Frame

The record must let another person reconstruct what happened:

- **The task.**
- **Relevant inputs.**
- **Source versions.**
- **Retrieval results.**
- **Configuration.**
- **Tool actions.**
- **Output.**
- **Changes.**
- **Reviewer response.**
- **Approval.**
- **Resulting system state.**

A model's narrative explanation of its answer is not proof of its internal path. Keep that narrative distinct from verifiable inputs, outputs, actions, and human reasons.

This is decision provenance rather than logging alone. The audit trail may show that a field changed at a particular time. Decision provenance connects that event to the evidence selected, the intermediate recommendation, the approval state, and the authority under which the change occurred. The organization does not need to preserve every internal computation a supplier cannot expose. It does need enough reliable evidence to explain its own regulated action and investigate a failure.

Segregation of duties also requires attention. If one agent retrieves the evidence, selects what is relevant, interprets it, drafts the conclusion, and evaluates its own result, several responsibilities that were separate in the human process may have collapsed into one technical pathway. A second agent is not independent merely because it has a different label. Independence depends on the data, methods, failure modes, and authority used for the check.

A complete audit trail can record a bad decision perfectly. The organization needs both a sound decision and reliable evidence of how it was made. Retention should follow the task, applicable predicate rules, risk, investigation needs, privacy, confidentiality, and contractual obligations. Indiscriminate logging can create uncontrolled copies without improving assurance. Exact reproducibility may be impossible for a probabilistic service, but the

organization should still preserve enough historical evidence to explain the regulated decision. If exact reproduction is required, that requirement should shape the system choice before approval.

Stage 3: Test the complete workflow

Stage 3 determines whether the configured system and assigned people can perform representative work under realistic conditions, first well enough for a bounded pilot and then well enough to support a separate routine-use decision.

Test the evidence and the review conditions

A citation is a starting point for verification. The source must exist, support the statement, and apply to the case. A true statement about another product, site, jurisdiction, process stage, or document revision can produce a wrong conclusion in the present workflow.

A summary can accurately describe five retrieved deviations and still omit a sixth that changes the conclusion. Citations to the five will not reveal the missing record. The team needs a justified method for identifying which records belong in the review and detecting consequential exclusions. Agreement is not an independent check when both reviewers depend on the same incomplete evidence path.

Make clear whether an output reports a fact, presents a calculation, offers an explanation, or proposes an action. Compare a factual statement with its source, repeat a calculation, and test an explanation against alternatives and contradictory evidence. Preserve the distinction between chronology and causation. If an excursion follows an intervention, the sequence may justify investigating a connection; it does not establish that the intervention caused the excursion.

The phrase "human in the loop" identifies a person's position, not the effectiveness of the review. Review depends on competence, evidence access, available time, interface design, and practical authority to reject or escalate. Showing a persuasive AI conclusion before the evidence may bias the reviewer's search. For consequential recommendations, the workflow can require an initial evidence assessment before displaying the AI conclusion, or another method that reduces anchoring. Lower-risk transformations may need only a direct source comparison.

Workload is part of the control. Increasing output while reducing the time available for each review can weaken the approval on which the process depends. A reviewer who may reject

an output but is penalized for delay receives conflicting instructions. The evaluation and later monitoring must examine review time, queue size, escalation patterns, and whether staff can report uncertainty and near misses without pressure to approve.

Treat automation bias as a control failure mode

For this method, automation bias is a foreseeable control failure mode when an AI-supported workflow shifts a reviewer from active analysis to passive monitoring. The risk increases when outputs are fluent, highly structured, usually correct, presented at high volume, or easier to accept than to challenge. Stage 3 must therefore test whether the review design can detect seeded and naturally occurring errors under realistic workload conditions rather than assume that the presence of a qualified reviewer provides detection. [33–39]

Forced verification steps. For consequential outputs, require the reviewer to examine the underlying evidence and record an independent finding before accepting or rejecting the AI recommendation. Identify the critical fields, claims, calculations, and source boundaries that must be checked. Make correction and escalation no more burdensome than acceptance, and test the control with seeded errors whose detection rate can be measured. A signature or checkbox without a recorded basis is not evidence that meaningful verification occurred.

Certainty label presentation. Treat objective verifiability and psychological risk as separate dimensions. A confidence score, traffic-light indicator, or high-certainty label must not conceal data gaps, assumptions, contradictory evidence, or the limits of the calculation. Compare matched review cases in which the indicator appears before and after the reviewer's initial evidence assessment. If early presentation reduces scrutiny, delay or withhold the indicator until the independent assessment is recorded. Certainty labels may prioritize review effort, but they must not eliminate periodic challenge testing.

Reviewer workload and rotation. Define review-volume, review-time, queue-size, and task-homogeneity thresholds before routine use. When a threshold is exceeded, rotate reviewers, narrow the queue, increase targeted sampling, add an independent examination for high-criticality cases, or pause the workflow. Monitor detection rates, unsupported acceptances, unnecessary overrides, review time, escalation patterns, and near misses by reviewer and period. Declining visible error rates should trigger analysis of both system improvement and possible loss of detectability.

The evaluation must preserve evidence identity. Record the test collection, document versions, inclusion rules, expected findings, tool configuration, reviewer instructions, scoring rubric, and any changes made after results were examined. Reusing the same visible cases for repeated tuning can turn a test into practice. Fresh cases or a protected holdout set are needed when the result will support authorization.

The comparison must count more than average accuracy. Track these metrics separately:

- **Critical misses.** Consequential failures that easier successes cannot offset.
- **False connections.** Unsupported relationships that could misdirect an investigation.
- **Correct findings rejected.** Sound findings rejected without an evidence-based reason.
- **Unnecessary escalations.** Incorrect referrals that add burden without improving control.

Measure the time required to reach an acceptable evidence packet, not merely the time to generate the first draft. Record withdrawals and incomplete work rather than silently excluding them. If the sample is small, present the uncertainty instead of converting a few cases into a precise-looking percentage.

Before evaluation, define the performance needed for the bounded use and the improvement that would justify the added intervention. A qualification method that raises examination scores without improving independently assessed work has not demonstrated its intended value. If an interface change produces the same benefit with less burden, that alternative deserves consideration. The added controls must earn their place through evidence.

Illustrative acceptance criterion: deviation chronology (Stage 3)

A critical omission is any excluded record that would change the investigator's conclusion. Authorization for the bounded pilot requires zero critical omissions across at least 59 challenge cases. The cases must be independent of one another, drawn to represent the authorized operating scope, including products, sites, record types, and terminology, and held out from any tuning or rehearsal.

Under those conditions, zero failures in 59 cases supports, with 95% one-sided confidence, a true critical-omission rate no higher than about 4.95%, using the exact binomial upper bound, $1 - 0.05^{1/59}$. The rule-of-three shortcut, $3/n$, gives 5.08%. Demonstrating a rate below 1% requires 299 zero-failure cases, usually rounded to 300. If one critical omission occurs, the same 5% claim requires at least 93 total cases with only that one failure, and the failure must be investigated before testing resumes. The team cannot add one passing case and treat the failure as offset.

If the cases are not independent or do not represent the authorized scope, these figures overstate the evidence. The organization must narrow the authorized scope to what the cases actually cover rather than extend the confidence statement beyond them. Non-critical errors, such as formatting or awkward wording, are tracked separately against a stated tolerance and cannot offset a critical miss.

Need to set defensible acceptance criteria for a specific workflow? Book a call about a Stage 0 and Stage 3 engagement: https://calendly.com/bdrapeau-gxpframe

Empirical research supports testing the combined arrangement. Vaccaro and colleagues' meta-analysis covered 106 experiments and 370 effect sizes. Human and AI combinations performed better than humans alone on average, but worse than whichever component performed best alone. The studies largely predate current generative systems, and none were run under GMP. They show why adding a reviewer cannot be assumed to improve an outcome. [22]

Qualify people for the role they actually perform

Awareness that AI can make mistakes is a beginning, not evidence of readiness. The test is whether someone can identify a consequential defect in representative work, explain why it matters, and accept, reject, correct, or escalate the result appropriately.

Challenge cases must include:

- **Correct outputs.** Confirm that reviewers do not reject acceptable work merely because AI produced it.
- **Awkward but sound outputs.** Separate presentation quality from evidentiary correctness.
- **Fluent dangerous outputs.** Test whether polished language conceals a consequential error or unsupported conclusion.
- **Incomplete evidence.** Determine whether reviewers detect a missing record that could change the decision.
- **Wrong-product sources.** Test whether apparently relevant evidence from another product is rejected.
- **Cases where blanket rejection is wrong.** Confirm that reviewers can accept sound work rather than treating universal rejection as safety.

Count both missed defects and unsupported accusations of error. Decide in advance which serious failures cannot be offset by success on easier items. Assessors must agree on what acceptable work looks like and why; unresolved scoring disagreement weakens the permission decision.

Dell'Acqua and colleagues' randomized study of 758 knowledge workers illustrates the importance of task boundaries. AI improved speed and completion within the tested capability range. On a deliberately selected task outside that range, the AI groups' correctness was

about 19 percentage points lower than the control group's. The study concerns consulting work with GPT-4, not pharmaceutical qualification. Its methodological lesson is still useful: evaluation should include conditions where assistance can mislead as well as conditions where it helps. [23]

Qualification must match responsibility. A person using a bounded tool to reformat approved text does not need the same demonstration as a quality approver evaluating an investigation or a system owner granting an agent write access. Training hours and course completion are inputs, not evidence that someone can perform a particular employer's work. This follows the readiness argument I developed for advanced-therapy manufacturing. [24]

A qualification program can also fail by blaming the operator for impossible conditions. If the interface hides source status, the evidence collection is incomplete, or volume makes careful review unrealistic, repeated retraining will not repair the workflow. Training cannot recover evidence the system never makes available.

A passing standard needs a reason. Define which outcomes matter, how cases represent the assigned work, and how assessors will distinguish an acceptable decision from a plausible but unsupported one. Use enough varied cases to avoid granting authorization on the strength of a rehearsed example. Where judgment is involved, calibrate assessors against shared examples and investigate material disagreement. If competent assessors cannot apply the standard consistently, the organization does not yet have a dependable basis for using the result to grant permission.

Qualification must be reassessed when its supporting conditions change. A reviewer qualified on one evidence repository, interface, language, and authority level must not automatically remain qualified after the system begins searching additional repositories or initiating workflows. Requalification can be targeted to the changed function. It need not repeat every prior activity, but the retained evidence must have an explicit rationale.

Challenge agents before routine use

An agent can turn one mistaken interpretation into changes across several systems. Problems can also arise when each tool works but the sequence fails. Consider an agent authorized to open a follow-up task and update a linked record. The first operation succeeds, the second times out, and the agent retries the entire sequence. It creates a duplicate task. The request and generated text were appropriate; the agent lost track of what had already happened.

The system must recognize repeated requests, check the state of prior operations, and expose incomplete work. Verify the actual records rather than treating the agent's completion message as proof. The test protocol must include:

- **Duplicate requests.**
- **Expired approvals.**
- **Unavailable sources.**
- **Changed records.**
- **Interrupted execution.**
- **Concurrent edits.**
- **Partial success.**

Recovery must distinguish reversible and irreversible actions. A record change may be reversed; an external notification cannot be unsent. Restoring one database may leave another system inconsistent. The recovery plan must specify reconciliation, compensating action, communication, and the authority to suspend the workflow.

Operating hours also matter. If an agent can take consequential action outside staffed periods, the organization needs a response arrangement for those periods. Otherwise, operation may need to be limited to times when responsible people can intervene. Training users to watch carefully cannot replace least privilege, state checks, observable execution, and tested recovery.

Stage 4: Authorize routine use and monitor

Stage 4 applies the gate distinction stated above. The process and quality owners must review pilot performance, critical failures, unresolved risks, review burden, recovery results, and operating controls before granting routine authorization. Any expansion requires another explicit decision.

Continued use requires determining whether the authorized conditions still apply. Model updates, retrieval changes, revised procedures, new products, staffing changes, and added permissions can alter performance or risk. McKinsey's learning-loop model in biopharma R&D offers a useful parallel: later evidence can inform earlier decisions while authority for protocol and regulatory changes remains human. It is a strategic model, not a GMP control standard. [25]

Monitoring must focus on consequential performance. Average quality can remain stable while a rare defect family worsens. Examine relevant products, sites, roles, document types,

languages, and failure classes where the data support those distinctions. Keep small samples and uncertainty visible.

Overrides and near misses can provide useful signals, but their meaning requires context. A sudden absence of overrides may reflect improvement, rubber-stamping, poor capture, or migration to an unapproved tool. Stage 0 inventory and ordinary user feedback help identify shadow use that formal metrics miss.

Connect workflow authorization to organizational management review

Every Gate Decision Record must contribute a defined set of fields to an organization-wide AI Authorization Register: the workflow and owners, current decision and authority level, acceptance criteria, review date, open risks, material changes, incidents and near misses, reviewer-performance signals, corrective-action status, and evidence location. The register is an index and aggregation layer, not a substitute for the controlled Gate Decision Record, validation evidence, deviation, change-control record, or CAPA. Aggregation must preserve distinctions by workflow, product, site, role, failure class, supplier, and authority level when a combined average could conceal a consequential problem. [8, 17–18]

A function sufficiently independent of the workflow's operation must periodically test whether authorization and monitoring controls work in practice. The review must use a justified, risk-based sample; trace register entries to source evidence; reconstruct why authorization was granted; compare authorized scope with actual use; examine rejected, withdrawn, and incomplete outputs; determine whether stop triggers were followed; and test whether corrective actions worked on fresh evidence. Independence depends on authority, incentives, access to evidence, and freedom from responsibility for the result, not on the reviewer's organizational label alone. The review may conclude that a control is effective, effective with conditions, partially effective, ineffective, or unable to be determined from the available evidence. [8, 17–18]

Material findings must enter the pharmaceutical quality system's management-review process with the affected workflows, evidence limits, required decision, responsible owner, due date, and method for checking effectiveness. Management action may change resources, policy, supplier controls, common technical controls, risk criteria, or the permitted class of use. Each decision must flow back to the affected workflow through the appropriate change control, CAPA, requalification, restriction, suspension, or retirement record. Management review does not replace quality-unit responsibilities or retroactively make an unsupported workflow acceptable. [8, 17–18]

Stage 5: Change, correct, reauthorize, and retire

Assess what a change could affect before selecting the tests. A wording edit that leaves behavior unchanged may need little additional assurance. A change in source selection may require targeted regression cases. Granting write access creates a different authority envelope and a different qualified state. The organization must document which prior evidence still applies and which conditions require new testing or requalification.

Approved changes must also reach the right people and systems. A revised procedure may affect particular sites, roles, work instructions, assessments, and permissions. Assigning the document to everyone is easy to count but does not show that the correct intervention occurred. Some changes require awareness; others require practice, observed demonstration, requalification, or a change in authorization. A governed applicability process must identify affected relationships from approved sources, show the basis for each proposed connection, and leave uncertain cases for responsible review. [26]

When a failure occurs, determine its scope, preserve the evidence, restrict affected use when warranted, and investigate human and system causes. Retraining is appropriate when a person lacked a required skill. It will not repair missing evidence, an interface that hides limitations, a defective permission design, or unreliable equipment. Corrective action must face fresh cases that reproduce the failure mechanism, along with unaffected cases that check for new errors.

Correctability means being able to find a problem, determine what caused it and what it affected, correct it under change control, and test whether the correction worked. Completing retraining or changing a prompt establishes that an action occurred, not that it worked. The final decision may be to reinstate the workflow, narrow it, stop it, or retire it. Retirement should address remaining records, access, supplier dependencies, and continuity obligations.

Count the time and cost of the whole job

AI saves little if people cannot check its work well enough to use it. Measure the time and effort required to finish an acceptable piece of work, including review, correction, escalation, maintenance, and downstream rework. Drafting speed alone does not show that the process improved. [27]

Sanofi reports reductions of up to 80% in time and effort for generative-AI-assisted manufacturing reports, beginning with annual Product Quality Reports (PQR) and extending to

other manufacturing documentation. The public account does not provide enough detail to calculate the cost of an approved report or the burden of all controls. It nevertheless provides a relevant operating example of realized drafting benefit. [28]

A published case study of AstraZeneca's AI-governance implementation describes approximately four full-time staff coordinating the effort during 2020 and 2021 and a 14-week audit requiring about 2,000 person-hours, without technical testing of individual models. These estimates belong to that program and do not measure reduced AI error. They show why governance labor must be included in the business case. [29]

Evidence outside pharma also shows why benefits should be measured by task and user. Brynjolfsson, Li, and Raymond's published study of 5,172 customer-support agents found a 15% average increase in issues resolved per hour, with larger gains among less-experienced workers. It was a staggered deployment in one firm, not a GMP trial. It cautions against assuming that the most experienced workers always benefit most or that AI necessarily prevents learning. [30]

Before a pilot starts, define what should improve and what cannot worsen. An SOP gap-check tool might be expected to reduce search time without increasing consequential omissions or false alarms. Measure the current process first. Compare time to acceptable work, important errors, unnecessary escalations, review burden, and maintenance cost over an agreed period.

Supplier agreements should identify the intended use, evidence each party will provide, acceptance criteria, change-notification duties, records available for investigation, and responsibility for evaluating changes. Delivery of software is not acceptance of the configured workflow. NIST's voluntary Generative AI Profile supports explicit supplier expectations for quality, security, provenance, third-party evaluation, and incident responsibilities, but it does not impose a pharmaceutical contracting rule. [31]

Pharmaceutical companies now have enough operating experience to identify recurring failure modes and propose controls. The public evidence base is not yet mature enough to establish the long-term effectiveness and net cost of a complete governance model. This approach is therefore designed as a testable operating method: it helps organizations decide what AI-supported work may proceed while generating evidence about whether the controls improve the work and justify their cost.

Putting the cycle to work in a fictional manufacturer scenario

Industry reporting describes the practical challenge of connecting inspection, audit, and deviation information across unstructured and structured records, including views by month, site, and authority; links to responses and after-action reviews; and translation, search-processing, and corpus-size tradeoffs. The following application is fictional. Its controls, scenarios, and acceptance decisions are proposals for illustrating the method, not descriptions of any organization's internal implementation or results. [32]

Consider this question: has a particular quality concern appeared elsewhere, and what evidence should a quality team examine before deciding whether wider action is warranted? The permitted output would be a cited evidence packet with unresolved questions. The system would not declare a site compliant, predict an inspection outcome, or approve corrective action. An authorized quality owner would decide the significance of the findings and any response.

Begin with a bounded evaluation collection reconciled against an inventory. Each consequential statement must link to its original passage, document status, site, date, and translation history where applicable. Include superseded responses and relevant records expressed in unfamiliar terminology. Outside a bounded collection, search limits should remain explicit because failure to retrieve a precedent does not establish that none exists.

Suppose the packet omits a decisive earlier finding recorded in unfamiliar terminology. The returned evidence looks consistent, and the reviewer concludes that the concern is isolated. Reconciliation against the test inventory reveals the omission. The evaluation owner preserves the query, retrieved records, configuration, and review decision, then restricts the affected use while the failure is investigated. This is an illustrative challenge, not a reported incident.

The investigation should distinguish at least three possibilities. If the record was never indexed or the search failed to match its terminology, correct ingestion or retrieval. If the interface concealed search limits or unavailable records, redesign the workflow. If the evidence and limitations were accessible but the reviewer misunderstood their significance, targeted preparation may be warranted. More than one cause may contribute.

Test the correction on fresh cases with similar retrieval problems and on unaffected cases that could reveal new errors. Measure consequential omissions, unsupported conclusions, correct packets wrongly rejected, unnecessary escalation, and review burden. Resolving the original case alone would not justify restoring use. The quality owner should decide whether

the new evidence supports reinstatement, narrower scope, continued restriction, or retirement.

For structured-data exploration, separate verifiable counts from interpretations. A challenge case can present an apparent rise in findings caused by a changed category definition, duplicate record, or different observation period. The calculation must be reproducible from the underlying dataset with filters and exclusions retained. The reviewer must distinguish a change in recorded data from evidence of worsening performance. When the data cannot support comparison across sites, the system must expose that limitation.

For connections across quality systems, test whether records describe the same issue, related issues, or only similar language. Pair an audit observation with a deviation from a different product and a superficially reassuring response that does not address the original concern. The reviewer must reject the unsupported connection, identify what remains unresolved, and avoid treating a completed response as proof of effectiveness.

Where feasible, compare three conditions: the current search-and-review process, AI assistance with existing preparation, and AI assistance after task-specific preparation. Keep the tool, settings, and source collection the same in the two AI groups so better software is not mistaken for better training. Use unfamiliar cases, distribute difficult cases fairly, and score the work without revealing the condition. Report baseline performance and uncertainty alongside any improvement.

How we would know the approach is not helping

The evaluation must permit the conclusion that an added check, workflow step, or training requirement does not help. A simpler interface, deterministic tool, narrower AI task, or existing process may achieve better results. An assessment that mainly produces documentation or teaches people to pass its own cases has not demonstrated workplace value.

The method is not helping if qualified reviewers cannot reproduce its decisions, if important omissions or false accusations do not improve, if the burden exceeds the task's benefit, if users move to unapproved tools, or if monitoring cannot detect deterioration soon enough to act. Those findings must lead to redesign, restriction, or withdrawal rather than another layer of paperwork.

Decide whether the work can proceed

Before allowing AI to support a consequential quality decision, name the task and accountable owner, examine representative results, and agree on what finding would make the team stop or restrict use. Set a date or trigger for reviewing the evidence again.

What evidence would justify authorizing this workflow? What finding would make us restrict or stop it? Answer both before the work proceeds.

GxP
Frame

Sources

1. Martha Heller. "How Roche drives AI outcomes in a changing world." CIO, September 16, 2026. Interview with Wafaa Mamilli. Read source
2. Boyd Spencer, Parag Patel, and Vivek Arora, with Krithiknath Tirupapuliyur and Raj Rajendran. "Gen AI: A game changer for biopharma operations." McKinsey, January 28, 2025. Read source
3. Eli Lilly and Company. "Responsible Use of AI." Corporate sustainability disclosure, updated July 1, 2026. Read source
4. GSK. Annual Report 2025. Manufacturing and supply section. Read source
5. US Food and Drug Administration. Purolea Cosmetics Lab, Warning Letter 320-26-58, MARCS-CMS 722591. April 2, 2026. Read source
6. 21 CFR 211.25, Personnel qualifications. eCFR text accessed September 18, 2026. Read source
7. International Council for Harmonisation. ICH Q9(R1), Quality Risk Management. Step 4 version, January 2025. Read source
8. International Council for Harmonisation. ICH Q10, Pharmaceutical Quality System. Step 4 version, June 2008. Read source
9. 21 CFR 211.68, Automatic, mechanical, and electronic equipment. Electronic Code of Federal Regulations. Read source
10. US Food and Drug Administration. Data Integrity and Compliance With Drug CGMP: Questions and Answers. Guidance for Industry, December 2018. Read source
11. Pharmaceutical Inspection Co-operation Scheme. PI 041-1, Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments. July 1, 2021. Read source
12. US Food and Drug Administration and European Medicines Agency. Guiding Principles of Good AI Practice in Drug Development. January 2026. Read source
13. European Commission. EudraLex Volume 4. Effective-document listing checked September 21, 2026. Read source
14. European Commission. Draft Annex 22, Artificial Intelligence, July 2025 consultation draft; consultation closed October 7, 2025. European Medicines Agency. GMP multistakeholder workshop on expert contributions to Annex 22 guidance development, June 30–July 1, 2026. Consultation · EMA workshop · Draft Annex 22
15. International Society for Pharmaceutical Engineering. ISPE GAMP Guide: Artificial Intelligence. July 2025. Read source
16. International Society for Pharmaceutical Engineering. GAMP 5: A Risk-Based Approach to Compliant GxP Computerized Systems. Second Edition, July 2022. Read source
17. Elham Tabassi. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, January 2023. Read source
18. International Organization for Standardization and International Electrotechnical Commission. ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system. December 2023. Read source
19. Brian Drapeau. "Europe Tried to Ban Generative AI From Critical GMP. The Ban May Not Survive. It Does Not Matter." Pharmaceutical Technology, August 13, 2026. Read source
20. Brian Drapeau. "Pharma's Next AI Problem Is Authority." Pharmaceutical Technology, August 27, 2026. Read source
21. US Food and Drug Administration. Part 11, Electronic Records; Electronic Signatures: Scope and Application. Guidance for Industry, August 2003. Read source
22. Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. "When combinations of humans and AI are useful: A systematic review and meta-analysis." Nature Human Behaviour 8, 2293-2303 (2024). Read source

23. Fabrizio Dell'Acqua and colleagues. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality." Organization Science, published online March 11, 2026. Read source
24. Brian Drapeau. "Training: The Real Constraint in CGT Manufacturing." Pharmaceutical Technology, May 27, 2026. Read source
25. Alex Devereson, David Champagne, Erika Stanzl, and Lieven Van der Veken, with Alex Peluffo, Chris Anagnostopoulos, and Jennifer Hou. "From linear gates to learning loops: Rewiring biopharma R&D with AI." McKinsey, June 29, 2026. Read source
26. Brian Drapeau. "Can Governance-by-Design Make AI Compliant in Pharma? (Part II)." Pharmaceutical Technology, July 16, 2026. Read source
27. Brian Drapeau. "Pharma's AI ROI Problem Isn't a Tech Problem but a Qualification Problem." Pharmaceutical Technology, August 10, 2026. Read source
28. Sanofi. "Digital & Artificial Intelligence." Manufacturing and supply section; accessed September 18, 2026. Read source
29. Jakob Mökander and Luciano Floridi. "Operationalising AI governance through ethics-based auditing: an industry case study." AI and Ethics 3, 451-468. Published online May 31, 2022; 2023 issue. Read source
30. Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. "Generative AI at Work." Quarterly Journal of Economics 140(2), 889-942 (2025). Read source
31. Chloe Autio and colleagues. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1, July 2024. Read source
32. Zachary Zubulake. "Lessons For AI Governance in Quality Systems." Pharmaceutical Technology, September 15, 2026. Conference reporting on Karthik Iyer's presentation. Read source
33. Frank Henrichmann and Oliver Herrmann. “Human-in-the-Loop as an Illusion of Control?” ISPE Pharmaceutical Engineering iSpeak Blog, September 4, 2026. Read source
34. Raja Parasuraman and Victor Riley. “Humans and Automation: Use, Misuse, Disuse, Abuse.” Human Factors 39(2), 230–253 (1997). Read source
35. Raja Parasuraman and Dietrich Manzey. “Complacency and Bias in Human Use of Automation: An Attentional Integration.” Human Factors 52(3), 381–410 (2010). Read source
36. J. Beck, S. Eckman, C. Kern, and F. Kreuter. “Bias in the Loop: How Humans Evaluate AI-Generated Suggestions.” Harvard Data Science Review 8(2) (2026). Read source
37. G. Romeo and D. Conti. “Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI.” AI & Society 41 (2026): 259–278. Read source
38. L. Kahn, E. Probasco, and R. Kinoshita. AI Safety and Automation Bias: The Downside of Human-in-the-Loop. Center for Security and Emerging Technology, Georgetown University, 2024. Read source
39. M. I. Saadeh, J. Janhonen, E. Beer, C. Castelyn, and D. N. Hoffman. “Automation Complacency: Risks of Abdicating Medical Decision Making.” AI and Ethics 5 (2025): 5783–5793. Read source

How to cite this paper

Drapeau, B. (2026). Governing AI in Pharmaceutical Operations: How to decide what AI may do and whether people can use it reliably (Version 4.0). GxP Frame. https://www.gxpframe.com/governing-ai/

An archival copy of Version 3.0 is available through Zenodo. This document is Version 4.0 and should be cited as such.

GxP
Frame

Version 4.0 changes

Reworked the paper's decision record, acceptance criterion, fictional worked example, and control-basis presentation.

Renumbered sources by first appearance, removed uncited sources, and replaced all public paper links with the Version 4.0 PDF.

Added a Stage 3 control subsection that treats automation bias as a systematic failure mode and specifies verification, certainty-label, workload, and reviewer-rotation requirements.

Added a Stage 4 organizational feedback loop connecting Gate Decision Records to an AI Authorization Register, independent effectiveness review, pharmaceutical quality-system management review, and controlled action back into affected workflows.

AI Work Authorization Toolkit

The toolkit page contains the reference inventory, Gate Decision Record, workflow scoring sheet, AI Authorization Register, Independent Effectiveness Review Record, user guide, and role cards that show the cycle applied to bounded fictional examples.

Author disclosure

The author offers training and a licensed toolkit based on this method.

This version does not name external reviewers or acknowledgments. A formal external-review process and any approved reviewer acknowledgments will be added in a subsequent version.

Brian Drapeau is the founder and principal consultant of GxP Frame and Life Sciences AI Academy. He has spent more than a decade building GMP training and quality systems across major biotechnology, pharmaceutical, and cell and gene therapy manufacturers, and writes on AI governance for Pharmaceutical Technology. LinkedIn · bdrapeau@gxpframe.com