Merivane

WHITE PAPER

# Efficient UAT for GxP Systems: Where Merivane.ai Helps

Risk-based User Acceptance Testing under 21 CFR Part 11 and EU GMP Annex 11, and how AI-assisted validation reduces effort without reducing assurance

# Table of contents


Executive summary ..... 3

1. The UAT problem ..... 3

2. What regulators actually require ..... 4

3. A modern, risk-based UAT model ..... 5

4. Where Merivane.ai helps ..... 7

5. Using AI in validation compliantly ..... 8

6. Expected benefits and adoption roadmap ..... 9

7. Conclusion ..... 10

References ..... 10

# Executive summary

User Acceptance Testing (UAT) is where a pharmaceutical company proves that a computerized system is fit for its intended use. It is also one of the most labor-intensive steps in computer system validation (CSV). Much of the effort goes into writing, formatting, reviewing and reconciling documents rather than into finding risk.

Regulators do not ask for that paperwork. 21 CFR Part 11, EU GMP Annex 11, GAMP 5 (2nd ed.) and FDA's 2025 Computer Software Assurance (CSA) guidance all ask for confidence, proportionate to risk, backed by trustworthy evidence. The gap between what is required and what companies actually produce is where efficiency lives.

Merivane.ai is an AI-driven validation and testing platform built for GxP environments. It targets the repetitive, document-heavy parts of UAT:

- turning user requirements into risk-assessed, traceable test cases;
- generating draft scripted and unscripted tests for human review;
- capturing execution evidence and audit-trailed results automatically;
- maintaining live traceability and release impact assessment for SaaS and agile systems;
- assembling UAT summary reports from the executed record.

Throughout, qualified people remain accountable: Merivane.ai drafts, links and checks; process owners and Quality review, approve and decide. That human-in-the-loop design is what allows AI to shorten UAT cycles while remaining consistent with Part 11, Annex 11 and the emerging EU Annex 22 on artificial intelligence.

# 1. The UAT problem

UAT in pharma is slow mainly because of how it is documented, not because of what it tests. The same patterns appear across QMS, LIMS, safety databases, eTMF, ERP and clinical systems.

| BOTTLENECK | WHAT IT LOOKS LIKE | CONSEQUENCE |
|---|---|---|
| Manual test authoring | Validation staff hand-write step-by-step scripts from the URS, often re-creating similar tests for every system and release | Weeks of authoring before execution starts; inconsistent quality |
| Uniform rigor | Every requirement tested with full scripts and screenshots, regardless of GxP impact | Effort spent on low-risk features; reviewers fatigued on high-risk ones |

| | | |
|---|---|---|
| **Spreadsheet traceability** | Requirements, risks, tests and results linked manually in Excel | Gaps found late, often by auditors |
| **Screenshot-heavy evidence** | Evidence pasted into Word, signed on paper or in a generic tool | Large packages, slow review, GDP errors |
| **Deviation handling by email** | Failures tracked outside the test record | Untraceable re-runs; data integrity risk |
| **SaaS release cadence** | Vendors ship multiple major releases a year | UAT never finishes; impact assessments rushed or skipped |
| **Report assembly** | Summary reports compiled by hand from executed packages | Days of rework at the end of every cycle |

The result is a validation function that becomes a bottleneck to digital programs, and a compliance record that is large but not necessarily strong. Inspectors repeatedly cite the same weaknesses: untested role permissions, audit trails "verified" by viewing a screen, re-runs without deviations, and supplier testing leveraged without rationale.

# 2. What regulators actually require

No regulation mandates a specific UAT format, screenshot count or script style. They require validated fitness for intended use, risk-based scaling and trustworthy records.

| SOURCE | STATUS (OCT 2026) | WHAT IT MEANS FOR UAT |
|---|---|---|
| **21 CFR Part 11 + 2003 Scope and Application guidance** | In force | Validate for accuracy, reliability and intended performance (§11.10(a)); verify access, audit trail, operational checks and e-signatures (§11.10(d)–(g), 11.50, 11.70, 11.200, 11.300) |
| **EU GMP Annex 11 (2011)** | In force | Risk-based validation (Cl. 1); URS traceability (4.4); evidence of test methods, scenarios, limits and error handling (4.7); data migration (4.8); supplier leverage (3) |
| **Draft revised Annex 11 + draft Annex 22 (AI)** | Consultation closed 7 Oct 2025; EudraLex still listed the 2011 Annex 11 and no Annex 22 as of Aug 2026 (ECA; Pharma Now) | Annex 11 grows to 17 chapters with heavier emphasis on security, identity and access, audit trail and cloud oversight; Annex 22 sets expectations for AI in critical GMP use |
| **FDA CSA guidance** | Final 24 Sep 2025; updated 3 Feb 2026 (NSF) | Select assurance activities by risk; unscripted testing and system-generated records (logs, audit trails) acceptable as evidence |

| | | |
|---|---|---|
| ISPE GAMP 5, 2nd ed. (2022) | Industry standard | Critical thinking, supplier leverage, scaled testing; explicit discussion of tools and automation |
| FDA / MHRA / PIC/S data integrity guidance | In force | ALCOA+ applies to the validation records themselves |

Two implications follow. First, efficiency is permitted: a company that tests less on low-risk functions, leverages supplier evidence and uses system-generated records is aligned with regulatory intent. Second, the controls must still be proven: Part 11 and Annex 11 data integrity controls in the company's own configuration must be challenged in UAT every time.

# 3. A modern, risk-based UAT model

An efficient UAT puts rigor where GxP risk is, leverages what suppliers have already proven, and always challenges the data integrity controls.

## 3.1 Scale testing to risk

| RISK TIER | EXAMPLE FUNCTIONS | TEST STYLE | EVIDENCE |
|---|---|---|---|
| High | Batch disposition, e-signature approval, calculations, audit trail, adverse event case processing | Scripted, with negative and boundary tests | System records or screenshots at critical steps; independent review |
| Medium | Workflow routing, GxP notifications, reports | Scenario-level scripted or objective-based unscripted | Scenario outcome and system records |
| Low | Cosmetic UI, non-GxP reference data | Unscripted, exploratory, or rely on supplier testing | Brief record of scope and outcome |

## 3.2 A core Part 11 / Annex 11 control test set

Every GxP system's UAT should challenge, in the company's configuration and with real role accounts:

- **Access and roles** (§11.10(d), (g); Annex 11 Cl. 12): each role can do exactly what it should; out-of-role actions are blocked.
- **Authentication** (§11.300): lockout, session timeout, no shared accounts.
- **Audit trail** (§11.10(e); Cl. 9): create, modify and delete actions captured with who, what, when, old/new values and reason; entries cannot be altered.

- **Electronic signatures** (§11.50, 11.70, 11.200; Cl. 14): name, date/time and meaning displayed; signature linked to the record; two-component signing enforced.
- **Workflow and data checks** (§11.10(f), (h); Cl. 5, 6): required sequence enforced; invalid inputs rejected; calculations independently verified.
- **Data transfer and copies** (§11.10(b); Cl. 4.8, 8): migrated and interfaced data match source; reports and exports complete.

## 3.3 Treat UAT records as GxP records

Test scripts, results, evidence and approvals must meet ALCOA+. Failures are raised as incidents and classified (script, tester, environment or system defect), never silently re-run. The summary report documents the release decision, with open defects risk-accepted by the process owner and Quality.

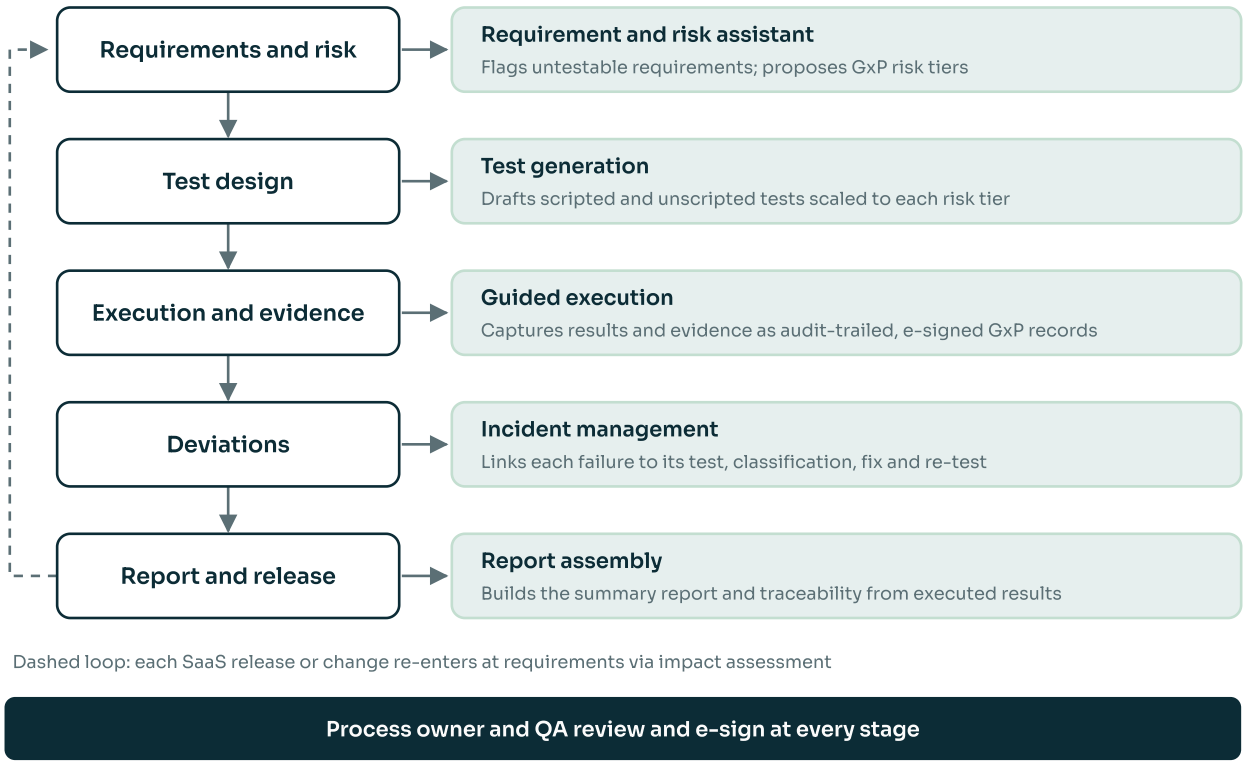
## 3.4 Build for continuous change

For SaaS and agile systems, UAT is no longer a one-time event. Each release needs an impact assessment and a targeted, repeatable regression suite covering high-risk workflows and the core control set.

# 4. Where Merivane.ai helps

Merivane.ai applies AI to each UAT stage where effort is high and judgment can stay with people: drafting, linking, checking and assembling. Decisions on risk, acceptance and release remain with the process owner and Quality.

### Merivane.ai supports every UAT stage; people approve each one




UAT lifecycle with Merivane.ai capabilities · 5 stages

The stages run top to bottom; Merivane.ai supports each one, and every output is reviewed and e-signed by people before it counts as validation evidence.

| UAT STAGE | TYPICAL MANUAL EFFORT TODAY | WHAT MERIVANE.AI DOES | WHAT PEOPLE STILL DO |
|---|---|---|---|
| Requirements and risk | Reading URS line by line; risk ratings in spreadsheets | Analyzes requirements for ambiguity and testability; proposes a GxP risk tier per requirement; builds the traceability skeleton | Confirm requirements and risk ratings; approve the risk assessment |
| Test design | Hand-writing scripts from scratch for every system and release | Drafts scripted tests for high-risk items, scenario or objective-based tests for medium and low risk | Review, edit and approve scripts; confirm expected results |

| | | | |
|---|---|---|---|
| **Execution and evidence** | Pasting screenshots into Word; wet-ink or generic e-signatures | Guides testers step by step; captures results and targeted evidence as audit-trailed records with Part 11-compliant e-signatures; applies automated GDP and completeness checks | Execute tests on the real system; record actual results; independent review |
| **Deviations** | Email and spreadsheet tracking | Raises incidents in context; supports classification; links fixes and re-tests to the original step | Investigate, classify and approve disposition; accept residual risk |
| **Report and release** | Days compiling summary reports and matrices | Generates the UAT summary report and final traceability directly from executed records; highlights open items | Review conclusions; make and sign the release decision |
| **Ongoing change (SaaS, agile)** | Rushed or skipped release impact assessments | Compares release notes and configuration changes to the requirement and risk baseline; proposes the targeted regression scope | Approve impact assessment and regression scope |

# 5. Using AI in validation compliantly

AI can accelerate validation only if the validation record stays trustworthy. Merivane.ai is designed around five principles that map to Part 11, Annex 11 and the direction of draft Annex 22.

| PRINCIPLE | HOW MERIVANE.AI APPLIES IT | REGULATORY ANCHOR |
|---|---|---|
| **Human accountability** | AI output is always a draft. Requirements, risk ratings, test scripts and reports are reviewed and approved by named, qualified people before use | Annex 11 Cl. 2 (personnel), Cl. 4; draft Annex 22 emphasis on human oversight; GAMP 5 critical thinking |
| **Transparency and traceability** | Every AI-generated artifact records its source inputs, model version and generation date, and is linked to its approver | ALCOA+ (attributable, original); Part 11 §11.10(e) |
| **Tool assurance proportionate to use** | Merivane.ai is a validation tool, not a GxP production system. Customers assess it for fitness for intended use, supported by supplier documentation, and verify outputs through human review rather than trusting the model | GAMP 5 (2nd ed.) treatment of tools; FDA CSA |
| **Part 11-compliant record keeping** | Unique accounts, role-based access, secure time-stamped audit trail, and electronic signatures with name, date/time and meaning on approvals and executions | Part 11 §11.10, 11.50, 11.70; Annex 11 Cl. 9, 12, 14 |

| | | |
|---|---|---|
| **Locked, controlled models** | The model version used to generate validation artifacts is controlled and recorded; model changes go through supplier change control and are communicated to customers | Annex 11 Cl. 10; draft Annex 22 focus on static, controlled models |

Two boundaries are deliberate. Merivane.ai does not decide whether a system is fit for use; it prepares the evidence for the people who do. And it does not replace testing of the customer's configuration: AI-generated scripts are executed against the real system by trained users or validated automation, and the results stand on their own.

# 6. Expected benefits and adoption roadmap

The value of AI-assisted UAT shows up in cycle time, review effort and inspection readiness. Each company should baseline its own metrics before adoption and measure against them in a pilot.

## 6.1 Benefits

| AREA | BENEFIT | METRIC TO BASELINE AND TRACK |
|---|---|---|
| **Test authoring** | Draft tests generated from requirements in hours instead of weeks of manual writing | Authoring hours per requirement |
| **Risk focus** | Effort concentrated on high-risk functions; low-risk testing can be scaled down | Share of tests by risk tier |
| **Review** | Smaller, structured evidence packages; automated completeness and GDP checks before QA review | QA review hours per package; review comments per package |
| **Traceability** | Live requirement-to-result coverage; gaps visible before execution | Untraced requirements at report stage |
| **Deviations** | Incidents captured in-tool and linked to tests | Deviations without documented disposition |
| **SaaS releases** | Faster impact assessment and targeted regression | Days from vendor release to validated use |
| **Inspection readiness** | Consistent, retrievable, audit-trailed validation records | Time to retrieve validation evidence on request; related findings |

### 6.2 Adoption roadmap

1. **Assess and qualify (weeks 0–4).**

   Supplier assessment of Merivane.ai, intended-use statement for the tool, and an SOP update describing AI-assisted test authoring with mandatory human review.

2. **Pilot (weeks 4–12).**

   Choose one medium-complexity GxP system or one SaaS release. Baseline current metrics, run UAT with Merivane.ai, and compare authoring hours, review effort and coverage.

3. **Standardize (months 3–6).**

   Build a reusable library: the Part 11 / Annex 11 core control test set, role-based access tests and common workflow scenarios. Integrate with requirements and change control.

4. **Scale (months 6–12).**

   Extend to the system portfolio and to SaaS release management; add automated regression where stable.

5. **Sustain.**

   Include AI-assisted validation in periodic review, internal audit and inspection readiness programs; monitor Annex 11 and Annex 22 finalization and update procedures accordingly.

## 7. Conclusion

Part 11 and Annex 11 ask pharmaceutical companies for confidence, not volume. Most of the cost of UAT today sits in manual authoring, uniform rigor, spreadsheet traceability and report assembly, none of which regulators require. Merivane.ai removes that overhead by drafting risk-based, traceable tests, capturing evidence as GxP records and keeping validated systems current through every release, while qualified people keep the decisions. The result is faster UAT, stronger evidence and a validation function ready for the revised Annex 11 and the arrival of Annex 22.

# References


1. 21 CFR Part 11, Electronic Records; Electronic Signatures (1997).
2. FDA, Part 11, Electronic Records; Electronic Signatures — Scope and Application (2003).
3. FDA, Computer Software Assurance for Production and Quality System Software, final guidance (Sept 2025; updated Feb 2026) — NSF summary.
4. FDA, Data Integrity and Compliance With Drug CGMP: Questions and Answers (2018).
5. EudraLex Vol. 4, EU GMP Annex 11: Computerised Systems (2011).
6. European Commission, draft revised Annex 11, draft Annex 22 (Artificial Intelligence) and draft Chapter 4, consultation July–October 2025 — ECA Academy overview.
7. Pharma Now, Annex 22 Is Not Law Yet (Aug 2026).
8. ISPE, GAMP 5: A Risk-Based Approach to Compliant GxP Computerized Systems, 2nd ed. (2022).
9. MHRA, GxP Data Integrity Guidance and Definitions (2018); PIC/S PI 041-1 (2021).

Merivane

merivane.ai