Validation Plan Outline: AI Agent Used in a GxP Process

Atlas Bio Systems · atlasbiosys.com · Outline version 1.0 · 19 October 2026

Status: Template outline. Each section names its source and that source's status (final, current, draft or industry good practice) as checked on 19 October 2026. Re-check draft status before relying on it.

General guidance, not legal or regulatory advice. Adapt every section to your own procedures; the validation plan, acceptance criteria and conclusion must be approved by your quality function. Free to copy and use inside your organization. Square-bracketed text is for you to replace.

1. Document control and approvals

Basis: Identify the plan and who approves it. Annex 11 section 2: defined responsibilities for process owner, system owner and others. Borrowed from draft Annex 22 (3.1, 4.2, 7.2): an SME approves the intended use and acceptance criteria, and the test plan is approved, before acceptance testing.

Sources: A11 (CURRENT EU GMP guidance); A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed)

Item	Your entry
Plan ID / version	[ID, version, date]
System owner / process owner	[names, roles]
Process subject matter expert	[name] - defines intended use and acceptance criteria
Quality approval	[name, role] - approves this plan before acceptance testing

2. System description

Basis: Describe the agent as a system: model, software, tools and the data it can reach. Annex 11 4.3 asks for an up-to-date description of critical systems, including data flows, interfaces and security measures.

Sources: A11 (CURRENT EU GMP guidance)

Item	Your entry
Agent name and purpose	[one sentence]
Model(s) and version(s)	[provider, model, version; how version changes are notified]
Instructions / system prompt	[document reference and version]
Tools, actions and permissions	[each tool the agent can call; read or write; which systems]
Data sources it retrieves from	[repositories, document sets, versions]
Hosting and supplier(s)	[where it runs; supplier agreements - Annex 11 section 3]

3. Intended use statement

Basis: Validate for a defined intended use, not the platform in general (FDA DI Q&A, question 3). CSA recommends examining the intended use of each feature, function or operation and documenting whether each is used as part of production or the quality system. Borrowed from draft Annex 22 3.1 (written for static, deterministic models; excludes LLMs): an SME-approved description, input characterisation and known limitations before acceptance testing.

Sources: DI (FINAL, drug CGMP); CSA (FINAL, non-binding; device production/QMS scope, context only for drug/biologic CGMP); A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed); FDAAI (DRAFT, not for implementation)

Item	Your entry
Process step and records touched	[e.g. deviation summary drafting in QMS; records: ...]

Permitted inputs	[sources, formats, common and rare variations]
Operations the agent performs	[list each operation separately]
Outputs and recipients	[what it produces; who receives it]
Out of scope / prohibited actions	[what the agent must never do]
Known limitations and problem inputs	[erroneous, ambiguous or biased inputs identified]

4. Risk assessment by operation

Basis: Annex 11 section 1: base the extent of validation on a justified, documented risk assessment covering patient safety, data integrity and product quality. CSA: an operation is high process risk when its failure may foreseeably compromise safety (a medical device risk); for drug/biologic CGMP, also rate patient safety, product quality and data integrity per Annex 11 section 1. Consider reasonably foreseeable failures, not only likely ones. Human review before use usually lowers CSA process risk. Draft Annex 22 says generative AI/LLMs should not be used in critical GMP applications (direct impact on patient safety, product quality or data integrity).

Sources: A11 (CURRENT EU GMP guidance); CSA (FINAL, non-binding; device production/QMS scope, context only for drug/biologic CGMP); A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed)

Item	Your entry
Operation	[from section 3]
Reasonably foreseeable failures	[e.g. plausible but wrong output; wrong source; action on wrong record; instructions hidden in input content]
Impact if it fails	[safety / quality / data integrity]
Existing controls	[downstream review, verification, monitoring, supplier evidence]
Risk rating and rationale	[high process risk / not high process risk, and why; data integrity impact]
Draft Annex 22 criticality	[critical / non-critical: direct impact on patient safety, product quality or data integrity? Draft Annex 22 says LLMs should not be used in critical GMP applications]

5. Test strategy (CSA-style, risk-based)

Basis: CSA names scripted testing (robust or limited) and unscripted testing (scenario, error guessing, exploratory, the last with high-level objectives and pass/fail criteria). Higher-risk operations may warrant scripted or hybrid testing. Annex 11 4.7: show appropriate test methods and scenarios, including data limits and error handling.

Sources: CSA (FINAL, non-binding; device production/QMS scope, context only for drug/biologic CGMP); A11 (CURRENT EU GMP guidance)

Item	Your entry
Scripted test cases	[known paths; prohibited actions refused; fields drafted correctly]
Unscripted testing	[error guessing / exploratory objectives: missing or conflicting sources, unusual records, injected instructions, task the agent cannot complete]
Credit taken for existing controls	[supplier evaluation, downstream verification, logs - CSA V.A.5]
Test environment and fixed configuration	[model version, instructions version, tools, data snapshot]

6. Acceptance criteria for variable outputs

Basis: Neither Annex 11 nor draft Annex 22 sets acceptance criteria for LLM-agent output; draft Annex 22 states it does not apply to generative AI or LLMs and that such models should not be used in critical GMP applications. Principles borrowed from draft Annex 22 sections 4 to 6: metrics for the intended use; criteria set and approved by an SME before testing; performance at least that of the process replaced; representative, sufficiently large, correctly labelled and independent test data. Repeated runs per case are a practical approach, not regulator language.

Sources: A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed); A11 (CURRENT EU GMP guidance)

Item	Your entry
Metrics	[what is measured for each operation]
Acceptance criteria	[threshold per metric; stricter criterion for critical errors; by input subgroup if needed]
Baseline of the current process	[how well the manual/current process performs today]
Test set	[size and rationale; coverage of input variations; how labels were verified; kept independent]
Repeat runs	[number of runs per test case; how results across runs are judged]

7. Human review points

Basis: Borrowed from draft Annex 22 3.3 and 10.5 (for models in the draft's scope): where the model gives input to a decision made by a person and testing effort has been reduced, state the operator's responsibility, monitor their training and performance, and keep records; depending on criticality and testing, this may mean reviewing every output. Electronic signatures belong to people: 21 CFR 11.100(a) requires each to be unique to one individual; Annex 11 section 14 expects the same impact as a hand-written signature and a permanent link to the record.

Sources: A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed); P11 (FINAL regulation, US); A11 (CURRENT EU GMP guidance)

Item	Your entry
Review point	[where in the workflow]
Reviewer role and training	[who; what they are trained to check]
What the reviewer sees	[agent sources, changes made, flags]
Authority to reject and how recorded	[procedure reference]

8. Traceability

Basis: Annex 11 4.4: user requirements traceable throughout the life cycle.

Sources: A11 (CURRENT EU GMP guidance)

Item	Your entry
Requirement -> risk -> test -> result	[matrix reference or tool]

9. Change control and revalidation triggers

Basis: Annex 11 section 10: changes, including configurations, under a defined procedure. Borrowed from draft Annex 22 10.1 to 10.4: evaluate every change for retest and justify any decision not to retest; configuration control; performance and input-drift monitoring. The trigger list below is suggested practice, not regulator text.

Sources: A11 (CURRENT EU GMP guidance); A22 (DRAFT, not adopted; covers static deterministic models and excludes generative AI/LLMs - principles borrowed)

Item	Your entry
Trigger: model or model version change	[assessor, retest scope, record]
Trigger: instructions / system prompt change	[assessor, retest scope, record]
Trigger: tool, permission or connected-system change	[assessor, retest scope, record]
Trigger: retrieved documents or data change	[assessor, retest scope, record]
Trigger: process or record change	[assessor, retest scope, record]
Trigger: input drift or performance drop	[monitoring metric and threshold]
Trigger: deviation or incident involving the agent	[link to incident process - Annex 11 section 13]

10. Periodic review and monitoring

Basis: Annex 11 section 11: periodic evaluation that the system remains valid. Draft revised Annex 11 14.1: review whether the system remains fit for intended use or needs full or partial re-validation.

Sources: A11 (CURRENT EU GMP guidance); A11R (DRAFT, not adopted)

Item	Your entry
Review frequency and rationale	[risk-based]
Monitored metrics	[performance, input drift, reviewer rejections, incidents]

11. Validation record and conclusion

Basis: CSA V.A.6: record intended use, risk analysis result, testing performed, issues and resolution, a conclusion on acceptability for the intended use, who tested and when, and review and approval. Prefer digital records such as logs and audit trails.

Sources: CSA (FINAL, non-binding; device production/QMS scope, context only for drug/biologic CGMP)

Item	Your entry
Summary of testing and issues	[reference]
Conclusion	[fit / not fit for the intended use in section 3]
Approval	[name, role, date]

Source register

Key	Source and status	URL
CSA	FDA, Computer Software Assurance for Production and Quality Management System Software. FINAL guidance, issued 3 February 2026 (CDRH, CBER). Scope: medical device production and quality system software; context only for drug/biologic CGMP.	https://www.fda.gov/media/188844/download
DI	FDA, Data Integrity and Compliance With Drug CGMP: Questions and Answers. FINAL guidance, December 2018.	https://www.fda.gov/media/119267/download
A11	EU GMP (EudraLex Vol. 4) Annex 11, Computerised Systems. CURRENT guidance, in operation since 30 June 2011.	https://health.ec.europa.eu/document/download/8d305550-dd22-4dad-8463-2ddb4a1345f1_en?filename=annex11_01-2011_en.pdf

A11R	EU GMP Annex 11, revised, consultation draft. DRAFT. Consultation 7 July to 7 October 2025; not adopted as of 19 October 2026.	https://health.ec.europa.eu/document/download/40231f18-e564-4043-94de-c031f813d38b_en?filename=mp_vol4_chap4_annex11_consultation_guideline_en.pdf
A22	EU GMP Annex 22, Artificial Intelligence, consultation draft. DRAFT. Consultation 7 July to 7 October 2025; not adopted as of 19 October 2026. Covers static, deterministic models in critical GMP use; states it does not apply to generative AI or LLMs.	https://health.ec.europa.eu/document/download/5f38a92d-bb8e-4264-8898-ea076e926db6_en?filename=mp_vol4_chap4_annex22_consultation_guideline_en.pdf
FDAAI	FDA, Considerations for the Use of AI To Support Regulatory Decision-Making for Drug and Biological Products. DRAFT guidance, January 2025, not for implementation.	https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
GAMP	ISPE GAMP Guide: Artificial Intelligence (public table of contents). INDUSTRY GOOD PRACTICE, published July 2025; not regulation. Paid guide.	https://ispewebassets.org/files/publications/guidance-documents/2025-TOC/GAMP_Guide-AI_TOC.pdf
P11	21 CFR Part 11, Electronic Records; Electronic Signatures. FINAL regulation, in force (binding, US).	https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11

Companion answer: https://atlasbiosys.com/answers/how-do-you-validate-an-ai-agent-used-in-a-gxp-process.html