

AI in GXP: Control Over Time Diagnostic Tool

Weighted maturity assessment with artifact mapping and FDA-oriented control checks

01

Scope the use case

Identify whether the AI use case sits in GMP, CCP, GLP, PV, a regulatory submission support context, or a medical device/software context. Some FDA expectations below apply broadly, while others are device specific.

02

Score each control

Use 0 to 4 for each item: 0 = not defined, 1 = partially defined, 2 = defined but inconsistently applied, 3 = controlled and consistently applied, 4 = optimized and continuously monitored.

03

Document evidence

For each line item, name the governing artifact that demonstrates control. Artifacts are evidence of control, not paperwork for its own sake.

04

Rate risk and act

Flag each item red, yellow, or green. Capture the specific gap and the action needed to close it.

05

Roll up the result

Complete the domain average and weighted contribution rows. If the Control Over Time domain averages below 3.0, classify the system as not in control regardless of the total score.

Scoring calculations:

- **Domain Average** = Sum of scores ÷ number of questions
- **Weighted Contribution** = Domain Average × Domain Weight
- **Overall Score** = Sum of all weighted contributions
- **R/Y/G Thresholds:** Red <2.0 | Yellow 2.0–2.9 | Green ≥3.0
- **Overall R/Y/G** based on Overall Score
- **Gating Rule:** If Control Over Time <3.0 → NOT in control

Regulatory Applicability Note

This diagnostic is designed for AI in the GxP life sciences industry broadly. The FDA checks in Appendix A distinguish between expectations that are broadly relevant to electronic records, data integrity, and risk-based assurance versus expectations that are more specific to AI-enabled medical devices or AI used to support drug and biologic regulatory decision-making.

Executive Summary

Summary Item	Assessment
System / use case	
Owner	
Date	
Overall weighted score (0–4)	
Maturity level	
Control status	
Top 3 risks	
Top 3 priority actions	

Maturity scale. 0.0–1.5 = Ad hoc | 1.6–2.5 = Defined | 2.6–3.2 = Controlled | 3.3–4.0 = Controlled over time

Gating rules. If the Control Over Time domain average is below 3.0, the system should be classified as not in control regardless of the overall weighted score.

1. Intended Use, Context of Use, and Risk Framing (ICH Q9 lens)

Weight: **15%**

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Is the intended use and specific decision the AI supports clearly defined?			Intended Use Statement; AI Use Case Description; User Requirements		
Is the context of use defined, including users, workflow, boundaries, and when the model should not be used?			Context of Use / Functional Specification; Process Map		
Is GxP impact classified and tied to patient safety, product quality, data integrity, or subject protection?			Impact Assessment; Risk Assessment		
Have foreseeable failure modes and harm scenarios been identified?			Risk Register; FMEA / Hazard Analysis		
Are acceptance criteria and risk acceptance thresholds defined before deployment?			Validation Plan; Risk Acceptance Memo		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 15%		

2. Data Integrity, Lineage, and Data Governance (ALCOA+ lens)

Weight: 20%

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Are source data, data owners, and data quality requirements defined?			Data Governance SOP; Data Source Inventory		
Is data lineage traceable from source through preprocessing, model input, output, review, and archival?			Data Flow Diagram; Audit Trail Design		
Are training, tuning, test, and production datasets segregated and controlled?			Dataset Register; Configuration Record		
Is dataset representativeness assessed for the intended population or process space?			Data Representativeness Assessment; Sampling Rationale		
Are data access, retention, backup, and audit-trail expectations defined for electronic records?			Part 11 Assessment; Security / Retention SOP		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 20%		

3. Model Credibility, Validation, and Assurance

Weight: 20%

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Is the model type, logic, input/output behavior, and known limitations documented?			Model Description; Architecture / Logic Summary		
Are performance metrics and statistical acceptance criteria aligned to the context of use?			Validation Plan; Statistical Analysis Plan		
Has the model been challenged with edge cases, stress cases, and known failure conditions?			Challenge Test Protocol; Test Evidence		
Is there an independent review or approval of validation evidence commensurate with risk?			Validation Report; QA Review / Approval		
Can the organization explain why the model is credible for this particular use case?			Credibility Assessment; Context-of-Use Justification		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 20%		

4. Deployment, Human Oversight, and Change Control

Weight: 20%

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Is the end-to-end workflow defined, including where AI output is reviewed, accepted, overridden, or escalated?			Workflow Map; SOP / Work Instruction		
Are user roles, training requirements, and decision rights defined?			RACI; Training Matrix		
Are human-in-the-loop expectations clear when the model influences GxP decisions?			Procedure; Review / Approval Matrix		
Is version control in place for model, code, prompts, configuration, and data dependencies?			Configuration Management Record; Version Log		
Is change control defined for updates, retraining, prompt changes, thresholds, and interfaces?			Change Control SOP; Release Record		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 20%		

5. Ongoing Monitoring, Deviations, and Lifecycle Review

Weight: 15%

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Are live performance metrics, drift indicators, and alert thresholds defined?			Monitoring Plan; KPI Dashboard		
Is periodic review defined to assess continued fitness for intended use?			Periodic Review SOP; Review Template		
Do adverse trends, anomalies, or performance degradation trigger investigation and CAPA?			Deviation / Investigation SOP; CAPA Procedure		
Is there a defined approach for revalidation after significant changes or degraded performance?			Revalidation Trigger Matrix; Validation SOP		
Are complaints, field issues, audit findings, or inspection observations fed back into model governance relevant?			Complaint Handling; Audit / Inspection Tracker		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 15%		

6. Control Over Time (Gating Domain)

Weight: 10%

Question / Control Check	Score (0-4)	Risk R/Y/G	Governing Artifacts	Gap Identified	Action Required
Is there a clearly assigned accountable owner for the AI system lifecycle?			System Owner Assignment; Governance Charter		
Can the organization detect when the model or workflow degrades over time?			Monitoring Dashboard; Trend Review		
Can the organization explain why performance changed and what changed?			Root Cause Analysis; Change History		
Can it act on that change in a controlled, documented way?			Change Control Record; CAPA / Action Log		
Can the team demonstrate the system remains in a validated or assured state over time?			Periodic Review Record; Continued Performance Evidence		
Domain Average Score			Average of question scores		
Weighted Contribution			Domain average × 10%		

Appendix A. FDA-Oriented Upgrades to Include in the Diagnostic

Use this appendix to tailor the diagnostic based on the actual regulatory context. These checks are recommended for upgrades from an FDA-oriented lens and should be added when applicable.

Area	Recommended upgrade item	Why it matters	Typical artifact
Drug / biologic regulatory submissions	Add an explicit Context of Use and Model Credibility section for any AI model used to generate evidence that supports safety, effectiveness, or quality decisions.	FDA's 2025 drug/biologic AI guidance is built around model credibility for a defined context of use.	Credibility Assessment; Model Risk Assessment; Validation Plan
AI-enabled medical devices	Add transparency, user information, and device-specific performance checks.	FDA's 2025 AI-enabled device guidance emphasizes lifecycle risk management, transparency, and performance across the total product lifecycle.	User Labeling; Human Factors / Usability Evidence; Performance Summary
Device modifications	Add a predetermined change control plan check where iterative updates are expected.	FDA has finalized PCCP guidance for AI-enabled device software functions.	PCCP; Change Protocol; Impact Assessment
Connected or cloud-based device/software	Add cybersecurity, patchability, and secure update controls.	FDA's cybersecurity guidance expects design, risk management, and documentation for cyber devices.	Cybersecurity Risk Assessment; SBOM / Patch Procedure; Secure Update SOP
Clinical investigations	Add electronic systems / electronic records / e-signature controls if the AI supports trial conduct or data capture.	FDA's 2024 Part 11 clinical investigations guidance focuses on trustworthy and reliable electronic systems and records.	Part 11 Assessment; e-System Validation / Assurance Record; Audit Trail Review
GMP manufacturing / QC	Add explicit data review, audit trail review, and investigation triggers for AI-generated outputs that affect batch or quality decisions.	FDA's data-integrity guidance and drug-manufacturing AI discussion emphasize reliable data and controlled use in manufacturing.	Audit Trail Review SOP; Exception Review; Investigation Trigger Matrix
Generative AI / prompt-based tools	Add prompt governance, output review criteria, approved use cases, and prohibited uses.	These tools can change behavior without classic model retraining; governance must still capture controlled intended use and review.	Prompt Library; Approved Use Matrix; Output Review SOP
Bias / subgroup performance	Add subgroup or process-segment performance checks wherever representativeness matters.	FDA has emphasized representativeness and credible evidence for the intended use.	Subgroup Performance Report; Data Representativeness Assessment

Appendix B. Quick Scoring Guide

Score	Interpretation	Practical meaning
0	Not defined	No control is defined or evidence is absent.
1	Partially defined	Some control exists, but scope, ownership, or evidence is incomplete.
2	Defined but inconsistent	The control exists on paper, but use is inconsistent or not sustained.
3	Controlled and consistent	The control is implemented, repeatable, and supported by evidence.
4	Optimized / continuously monitored	The control is effective, trended, and improved through lifecycle review.

Appendix C. Worked Scoring Example

Use case: AI model supporting deviation triage in GMP manufacturing.

Question	Score (0-4)
Data sources defined and controlled	3
Data lineage traceable end-to-end	2
Training vs production data segregated	3
Data representativeness assessed	2
Part 11 / audit trail controls defined	3

Domain Average = $(3 + 2 + 3 + 2 + 3) \div 5 = 2.6$

Weighted Contribution (Data Integrity, 20%) = $2.6 \times 0.20 = 0.52$

R/Y/G = Yellow (2.0-2.9 → defined but inconsistent)

Domain	Avg Score	Weight	Weighted Contribution	R/Y/G
Intended Use & Risk	3.2	15%	0.48	Green
Data Integrity	2.6	20%	0.52	Yellow
Model Validation	2.8	20%	0.56	Yellow
Deployment Control	3.0	20%	0.60	Green
Monitoring	2.4	15%	0.36	Yellow
Control Over Time	2.5	10%	0.25	Yellow

Overall Score = $0.48 + 0.52 + 0.56 + 0.60 + 0.36 + 0.25 = 2.77$

Gating Rule: Control Over Time = 2.5 (<3.0) → NOT IN CONTROL

Interpretation: System appears controlled at deployment but not across lifecycle. Priority actions should focus on monitoring, drift detection, and lifecycle governance.