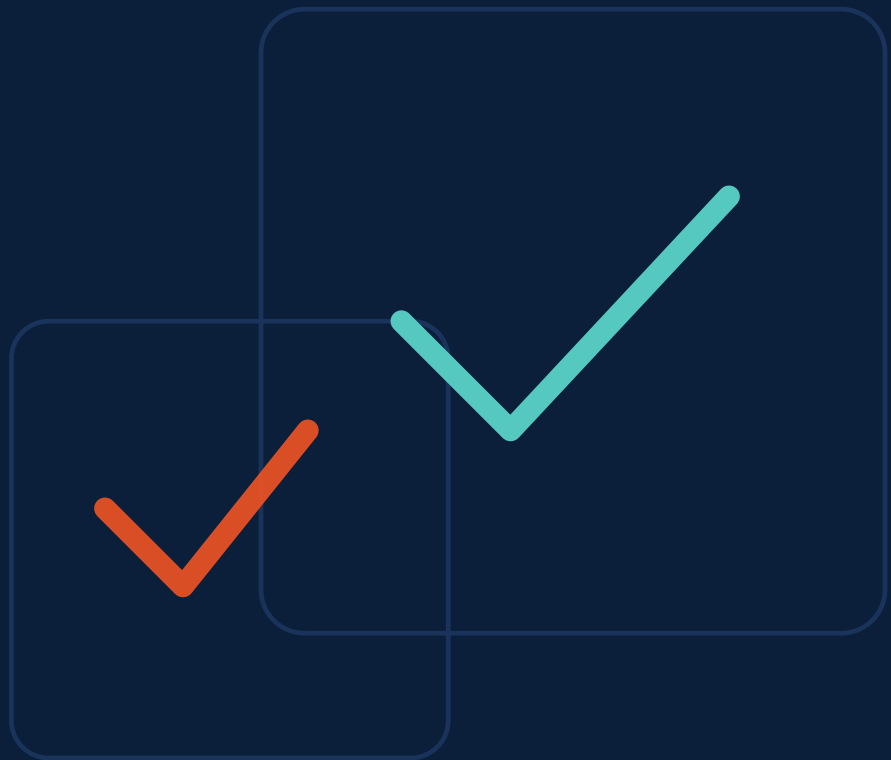


EBOOK · GOVERNANCE BENCHMARK SERIES



Audit-Ready by Design.

A Data Governance Benchmark for AI in Regulated Life
Sciences



Five pillars, four maturity levels — score your organization
against what regulators actually inspect.

kasmoglobal.com

The Stakes: Regulators Are Moving Faster Than Most Governance Programs

AI in life sciences no longer needs an adoption pitch. It needs a governance program regulators will actually accept — and the bar is being set faster than most internal programs can keep up with.

500+

FDA's Center for Drug Evaluation and Research reviewed **more than 500 drug submissions containing an AI/ML component** between 2016 and 2023 — and the pace is accelerating.¹

In January 2025, FDA published draft guidance on using AI to support regulatory decision-making for drug and biological products. A year later, FDA and the European Medicines Agency jointly published *Guiding Principles of Good AI Practice in Drug Development* — ten shared principles both agencies now expect sponsors to follow.¹ FDA's own CDER AI Council, formed in 2024, exists specifically to keep that bar consistent across every submission it reviews.¹

Then FDA did something most sponsors didn't expect: it started using agentic AI itself. ELSA now supports internal drug protocol reviews and CDRH-GPT supports device evaluation — both built around the same standard the agency is asking industry to meet: traceable edits and signed audit trails on every output.² As one 2026 industry analysis put it, auditability has evolved from a quality metric into a fundamental requirement.²

45%

of medtech leaders say agentic AI is already in production somewhere in their organization.

7%

describe that AI as fully embedded and audit-ready.³

What Inspectors Actually Ask

Most governance conversations start with frameworks. Inspectors start with three questions — and if you can't answer all three about a specific AI system, the framework doesn't matter yet.⁴

- 1 **Can you reconstruct every decision?** Full traceability showing which data inputs, which model version, and which human validated the output.
- 2 **Can you prove performance stays in range?** Documented acceptance criteria, drift monitoring, and confidence scoring tracked over time — not a one-time validation report.
- 3 **Can you show continuous compliance?** From initial validation through ongoing monitoring to system retirement, on demand, not reconstructed after the inspection is announced.

Regulators apply the same data-integrity discipline to AI outputs that they've always applied to GxP records — ALCOA++: attributable, legible, contemporaneous, original and accurate, plus complete, consistent, enduring and available. In practice, that means every AI-assisted decision needs a timestamped audit entry showing the model version, the decision, its confidence score, and the reviewer's final call.⁴

Three gaps inspectors specifically target:

MISSING MASTER INVENTORY

No centralized list of which AI systems exist, their risk classification, or validation status.

VENDOR BLIND SPOTS

Contracts that don't guarantee access to a third-party AI vendor's documentation during an audit.

GUARDRAIL GAPS

Language models deployed without confidence scoring or mandatory human review of low-confidence output.

⁴ Industry pharmacovigilance inspection guidance, 2026 — questions and gaps reflect current GxP/ALCOA++ inspection practice for AI-assisted systems.

Why Governance Is the Real AI Risk

The industry's instinct is to worry about whether a model is smart enough. The evidence says worry about whether it's governed enough.

In controlled testing with clinically designed scenarios, leading language models repeated or elaborated on a planted factual error in **up to 83% of cases**.⁵ Retrieval-augmented generation — anchoring a model's answer to verified source documents instead of letting it reason freely — cut that hallucination rate by **60–80%**.⁵

83%

of clinically-designed test cases saw a leading LLM repeat or elaborate on a planted factual error — the gap RAG-grounded, governed architecture is built to close.⁵

The difference between those two numbers isn't model quality. It's architecture: whether the system was built to cite its sources, or built to sound confident. That's the same distinction we drew in Kasmó's first e-book, *The Governed Agent*: a 90%-accurate agent making a clinical or regulatory call isn't a success, it's a liability. Governance is what turns a capable model into a defensible one.

“Success depends on clarity, provenance and disciplined human oversight — not model novelty.”²

⁵ Controlled LLM hallucination testing using clinically designed vignettes, 2026 pharmacovigilance AI governance research.

The Benchmark: Five Pillars of Audit-Ready AI Governance

We built this benchmark from what regulators are actually asking for — FDA and EMA's joint AI principles, current inspection practice, and the architecture we build into every Kasmo solution. Score your organization honestly against each pillar before you read any further.

1

Data & Model Lineage

Can you trace every output back to its source data, document and model version?

2

Human-in-the-Loop & Validation

Are review rates, ramp-down criteria and validation benchmarks defined and documented?

3

Access Control & Security

Is access role-based, logged and segregated by function, end to end?

4

Model Risk & Explainability

Are bias testing, drift monitoring and confidence scoring built in, not bolted on?

5

Documentation & Vendor Governance

Does a master AI inventory exist, and can every vendor contract survive an audit?

Turn the page for the full scorecard.

The Scorecard

Circle where your organization sits today on each row. Score each pillar 1–4, then add them up.

PILLAR	1 · AD HOC	2 · EMERGING	3 · MANAGED	4 · AUDIT-READY
Data & Model Lineage	No consistent link between an output and its source data or model version.	Lineage exists for a few priority systems; mostly manual reconstruction.	Lineage is documented and queryable for most production AI systems.	Every output traces automatically to source data, document and model version, on demand.
Human-in-the-Loop & Validation	Review happens informally; no defined rate or criteria.	Review rates set for some systems; ramp-down criteria undocumented.	Review rates and ramp-down criteria documented for priority systems.	Every system has defined review rates, documented ramp-down, and validated benchmarks.
Access Control & Security	Shared credentials; access not tied to role or logged.	Role-based access on some systems; inconsistent audit logging.	Role-based access and audit logs standard across most AI systems.	Enterprise-wide RBAC, full audit trail, and segregation of duties by design.
Model Risk & Explainability	No bias testing or drift monitoring in place.	One-time validation; no ongoing drift or bias monitoring.	Drift monitoring and periodic bias testing on priority systems.	Continuous drift/bias monitoring and confidence scoring on every production system.
Documentation & Vendor Governance	No inventory of AI systems in use; vendor contracts undefined on AI terms.	Partial inventory; vendor audit rights inconsistent.	Master inventory maintained for most systems; standard vendor audit clauses.	Complete, current master inventory; every vendor contract guarantees audit access.

■ 1 pt
 ■ 2 pts
 ■ 3 pts
 ■ 4 pts

5–9

Ad hoc. Every AI deployment right now is inspection risk.

10–13

Emerging. Pilots are running ahead of governance.

14–17

Managed, but gaps will surface under inspection.

18–20

Audit-ready. Extend this architecture, don't rebuild it.

Pillar Deep Dive: Lineage & Human Oversight

1 Data & Model Lineage

WHAT REGULATORS EXPECT

A timestamped audit entry for every AI-assisted decision, showing the model version, the data referenced, and the confidence score behind it — reconstructable on demand, not after the fact.⁴

HOW KASMO BUILDS IT IN

Unity Catalog governs every Kasmo solution end to end, so lineage isn't a report you generate before an audit — it's a property of the platform every output already carries.

2 Human-in-the-Loop & Validation

WHAT REGULATORS EXPECT

Defined initial review rates (often 100% during validation), documented ramp-down criteria tied to measured performance, and validation against expert or gold-standard benchmarks — not just efficiency targets.⁴

HOW KASMO BUILDS IT IN

Every Kasmo agent is advisory by design. The Clinical Trial Intelligence Agent and R&D Agents both route through a mandatory review workflow before any output reaches a real decision — a control, not an afterthought.

⁴ Industry pharmacovigilance inspection guidance, 2026.

Pillar Deep Dive: Access, Risk & Documentation

3 Access Control & Security

WHAT REGULATORS EXPECT

Role-based access, complete audit logs, and segregation of duties across every system touching regulated data — the same standard applied to any GxP system of record.

HOW KASMO BUILDS IT IN

Unity Catalog's access controls and audit logging apply uniformly across Commercial Copilot, the Clinical Trial Intelligence Agent, R&D Agents and the Supply Chain platform — one governance layer, not four.

4 Model Risk & Explainability

WHAT REGULATORS EXPECT

Bias testing across demographic subgroups, real-time drift monitoring, and confidence scoring on every output — validated before go-live and monitored continuously after.⁴

HOW KASMO BUILDS IT IN

MLflow tracks and versions every scoring model Kasmo ships — from HCP engagement ranking to site-selection scoring — so every score is explainable and auditable, not a black box.

5 Documentation & Vendor Governance

WHAT REGULATORS EXPECT

A master inventory of every AI system — purpose, risk classification, validation status, responsible owner — plus vendor contracts that guarantee audit access to algorithm documentation.⁴

HOW KASMO BUILDS IT IN

Each Kasmo solution ships with its own documented architecture and review workflow, and deploys inside your own AWS/Databricks workspace — so your data and your audit trail never leave your governance boundary.

⁴ Industry pharmacovigilance inspection guidance, 2026.

Kasmo's Governance Architecture, Mapped to the Benchmark

Governance isn't a module we add before a client's next inspection. It's the same foundation underneath every Kasmo solution — which is why the mapping below looks the same regardless of which pillar you scored lowest.

BENCHMARK PILLAR	KASMO ARCHITECTURE COMPONENT
Data & Model Lineage	Unity Catalog — automatic lineage from every agent output back to its source dataset, document or manufacturing event.
Human-in-the-Loop & Validation	Mandatory review workflows built into every agent — advisory by design, with qualified humans holding final sign-off.
Access Control & Security	Role-based access, audit logging and segregation of duties, applied uniformly across all four solutions.
Model Risk & Explainability	MLflow-tracked, versioned scoring models with documented, explainable logic behind every ranking or recommendation.
Documentation & Vendor Governance	Solution-level architecture documentation, deployed inside your own AWS/Databricks workspace and cloud of choice.

This is also why Kasmo's approach differs from a generic AI platform retrofitted to life sciences: the governance layer isn't a configuration option you turn on later. It's the same architecture whether you're scoring a 6 today or an 18.

How to Use This Benchmark

Your total score points to a different next step — not a different level of urgency. Every score below 18 is inspection risk sitting somewhere in your AI portfolio today.

Scored 5–13?

Start with a governance gap assessment before scaling any additional agentic AI use case. Pilots are outrunning your ability to defend them.

Scored 14–17?

Close the specific gaps this scorecard surfaced — likely documentation, vendor terms, or drift monitoring — before your next inspection cycle, not after a finding.

Scored 18–20?

Extend the same governed architecture to your next use case instead of standing up a separate one. You've already built the hard part.

Get Your Governance Benchmark Session

Bring your scorecard. We'll walk through where the gaps actually are, and what closing them looks like on your own AWS/Databricks environment.

[TALK TO KASMO GLOBAL](#)

Governance Is the Product Now

Regulators aren't waiting for the industry to catch up — they're already running agentic AI internally and building their own expectations around what they build. The sponsors and manufacturers who treat governance as core architecture, not a compliance checkbox added at the end, are the ones whose AI survives its first real inspection.

None of the five pillars in this benchmark are new regulatory demands. They're the same data-integrity discipline life sciences has applied to every GxP system for decades — now applied to a system that reasons and acts instead of one that just stores records.

The difference between an AI pilot that scales and one that gets shut down after a single inspection finding isn't the model. It's whether governance was the architecture from day one.

