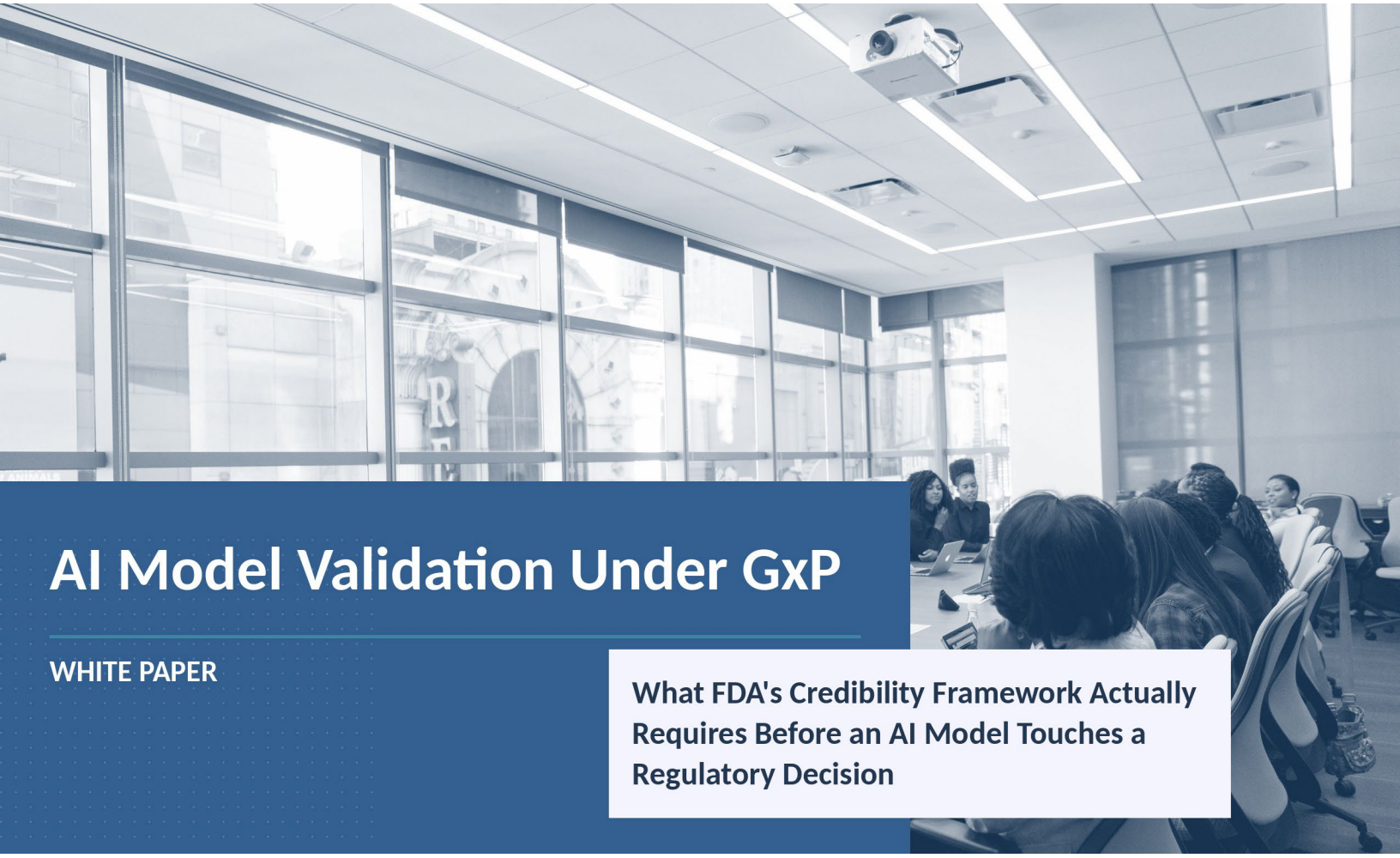
OMNITEK CONSULTING



AI Model Validation Under GxP

WHITE PAPER

What FDA's Credibility Framework Actually Requires Before an AI Model Touches a Regulatory Decision

Abstract

FDA has now reviewed more than 500 drug and biological product submissions containing AI components since 2016, and that volume has climbed sharply as AI tooling matured through 2025. The agency's January 2025 draft guidance responded with a seven-step, risk-based credibility framework, and in January 2026, FDA and EMA jointly released Guiding Principles of Good AI Practice in Drug Development, aligning both agencies around proportional validation, defined context of use, and lifecycle performance monitoring. An estimated 69% of pharmaceutical organizations now invest in AI, and roughly a third are already using it in process development or quality control, which means most life sciences organizations are validating AI models under a still-forming regulatory framework rather than a settled one. Traditional computer system validation was never built for a model whose behavior can shift as it encounters new data. This white paper applies established computer system validation and PMO governance best practices to validating AI models under GxP that satisfy where the regulatory guidance is heading, not just where it currently stands.

A Validation Discipline Built for the Wrong Kind of System

Computer system validation, as most quality organizations practice it, assumes a system behaves the same way every time it is given the same input. Validate it once against a defined set of test cases, document the results, and the system's behavior is established for the life of that validation, subject to a defined change control process if anything gets modified. That assumption holds reasonably well for a deterministic system. It breaks down for an AI model, particularly one that continues learning or whose performance can drift as the data it encounters in production differs from the data it was trained and validated on.

Quality and validation teams that try to force an AI model through a traditional computer system validation process often end up with documentation that satisfies the letter of the process while missing what actually matters: whether the model's specific behavior, in its specific context of use, can be trusted for the specific regulatory decision it is meant to support. FDA's own numbers show why this gap now carries real regulatory weight. The agency has reviewed more than 500 submissions with AI components since 2016, and that volume is not slowing down.[i]

What the Regulatory Framework Actually Asks For

FDA's January 2025 draft guidance lays out a seven-step credibility framework: define the question of interest, define the context of use, assess model risk, develop a credibility assessment plan, execute the plan, document results and deviations, and determine model adequacy for the defined context of use.[i] The framework is explicitly risk-based, meaning the rigor applied to any given model scales with how consequential its output is to the regulatory decision it supports, not a single fixed validation standard applied uniformly to every model regardless of stakes.

The framework's emphasis on context of use is the detail most organizations underestimate. A model is not validated in the abstract. It is validated for a specific question it is being asked to help answer, and a model considered adequately credible for one context of use may not be credible for a different one, even using the identical underlying algorithm. Organizations that validate a model once and then apply it broadly across multiple use cases are extending a credibility finding beyond the scope it was actually established for.

In January 2026, FDA and EMA moved from separate national guidance toward genuine alignment, jointly releasing Guiding Principles of Good AI Practice in Drug Development. The joint principles emphasize a human-centric, risk-based approach, proportional validation, robust data governance, multidisciplinary expertise, transparent model development practices, lifecycle performance monitoring, and clear communication of AI limitations, spanning use from early research through clinical trials, manufacturing, and post-market safety surveillance.[ii] For organizations operating across both U.S. and European markets, that alignment is significant. It means a validation approach built to satisfy one agency's expectations is now far more likely to also satisfy the other's, reducing the risk of building two separate, incompatible validation practices.

Why Most Organizations Are Behind Without Realizing It

An estimated 69% of pharmaceutical organizations now invest in AI, and roughly a third report already using it in process development or quality control, with many planning submissions involving AI in drug manufacturing within

the next five years.[ii] That adoption curve has moved faster than most organizations' validation practices have matured to match it. A team that adapted its computer system validation SOP to loosely accommodate AI, rather than building a genuinely fit-for-purpose credibility assessment practice, may pass an internal audit today while still being significantly under-prepared for what regulatory review actually expects.

The gap tends to be invisible until a specific submission or inspection forces it into view, because the shortfall is not usually a documentation failure regulators can spot immediately. It is a credibility assessment that never rigorously interrogated whether the model's validated performance actually generalizes to its intended context of use, buried inside documentation that otherwise looks complete.

Applying Best Practices: A Credibility-Based Validation Practice

Treating Technology and Data Enablement as a core PMO maturity dimension is an established governance best practice, and AI model validation is where that dimension now faces its hardest test. Applying that discipline as a genuine credibility-based validation practice means building around the regulatory framework directly, rather than retrofitting a traditional CSV process.

1. **Define Context of Use Before Validation Begins.** Document precisely what question the model is being asked to help answer and in what regulatory context, since a credibility finding is only as broad as the context of use it was actually established for.
2. **Scale Validation Rigor to Actual Risk.** Apply the risk-based principle genuinely: a model informing an exploratory internal analysis warrants a different validation depth than one whose output feeds directly into a regulatory submission or a GMP release decision.
3. **Build Lifecycle Monitoring Into the Validation Plan From the Start.** A model's credibility assessment cannot be a one-time event if the model's behavior can drift. Define what ongoing performance monitoring looks like and what triggers re-validation before the model goes live, not after a problem surfaces.
4. **Document Deviations as Rigorously as Successes.** The credibility assessment plan's requirement to document results and deviations means capturing where the model underperformed or behaved unexpectedly during validation, not just the results that support the intended conclusion.

None of this requires abandoning existing computer system validation infrastructure. It requires layering a genuine, context-specific credibility assessment on top of it, built to the standard the FDA and EMA framework actually describes rather than a lighter internal adaptation of a process designed for deterministic systems.

What Credibility-Based Validation Buys, and Where It Gets Hard

Advantages

- **Genuine regulatory readiness:** a validation practice built directly around the FDA and EMA credibility framework is positioned for the standard regulators are actually applying, rather than a retrofitted approximation of it.
- **Alignment across major markets:** building to the joint FDA and EMA principles reduces the risk of maintaining two incompatible validation practices for organizations operating in both regions.

- **Earlier detection of model drift:** lifecycle monitoring built into the validation plan from the start catches performance degradation before it becomes a regulatory or quality finding.
- **Defensible scope for model reuse:** explicit context-of-use documentation gives quality teams a clear basis for deciding whether an existing model's validation actually extends to a new proposed use, rather than assuming it does.

Risks and Barriers

- **Genuinely new expertise requirements:** credibility assessment as FDA and EMA describe it requires multidisciplinary expertise spanning data science, quality, and regulatory affairs that many validation teams do not yet have in-house.
- **A still-evolving regulatory target:** FDA's guidance remains in draft form and the joint principles are explicitly described as a foundation rather than a fixed standard, meaning validation practices need to be built for adaptability rather than permanence.
- **Cost of proportional but genuine rigor:** risk-based validation still requires real assessment work for higher-risk models, and organizations accustomed to a lighter validation touch may underestimate the resourcing this requires.
- **Retrofitting existing models:** models already in production without a genuine credibility assessment behind them represent a real remediation project, not just a documentation update.

Case in Point: The Model Validated for One Use and Deployed for Another

A mid-size biotech validated a machine learning model to flag potential data anomalies in a specific early-phase clinical trial's laboratory results, establishing solid credibility evidence for that narrow context of use. Encouraged by the model's performance, a different program team adopted the same model, unmodified, to screen laboratory data across a much larger late-phase trial with a materially different patient population and testing protocol, assuming the original validation extended to the new use.

During an internal quality review ahead of a planned regulatory submission, the team could not produce credibility evidence specific to the new context of use, only the original narrow validation. Establishing genuine credibility evidence for the actual deployed context required a dedicated assessment that delayed the submission timeline by several weeks. A validation practice that had documented context of use explicitly from the start, and required a fresh credibility assessment for any materially different application, would have caught the gap before the model was deployed to the second trial rather than during a pre-submission review.

Recommendations for Quality and Data Science Leaders

1. **Inventory every AI model currently in use against its documented context of use.** and flag any model deployed beyond the scope its original validation actually covers.
2. **Build a genuine risk-based tiering system for AI model validation.** so validation rigor scales deliberately with actual regulatory and quality consequence rather than applying a single standard uniformly.

3. **Define lifecycle monitoring requirements at the validation planning stage.** not as an afterthought once the model is already in production.
4. **Build cross-functional review into every credibility assessment.** combining data science, quality, and regulatory affairs expertise as the joint FDA and EMA principles specifically call for.
5. **Treat context-of-use changes as triggering a new credibility assessment.** rather than assuming an existing validation automatically extends to a new application.
6. **Monitor the evolving regulatory guidance directly.** since FDA's framework remains in draft form and the joint principles are explicitly designed to evolve with the technology.

Measuring Success: KPIs for AI Model Validation Maturity

- **Context-of-use documentation coverage:** the percentage of production AI models with an explicitly documented, current context of use on file.
- **Risk-tiered validation completion rate:** the percentage of models with a validation depth matched to their assigned risk tier, rather than a uniform default level.
- **Lifecycle monitoring coverage:** the percentage of production models with active, defined performance monitoring against drift or degradation.
- **Model reuse review rate:** the percentage of instances where an existing model was proposed for a new context of use that received a fresh credibility assessment before deployment.

Conclusion

FDA's 500-plus AI-inclusive submissions since 2016, its January 2025 credibility framework, and the January 2026 joint principles with EMA together describe a regulatory environment that has moved well past treating AI validation as an edge case. With an estimated 69% of pharmaceutical organizations now investing in AI, the organizations still validating these models through a lightly adapted traditional computer system validation process are working against a standard regulators have already moved beyond.

Applying disciplined best practices to a genuine credibility-based validation practice means building around context of use, proportional risk-based rigor, and lifecycle monitoring from the start, not retrofitting them in after a submission or inspection reveals the gap. Organizations that make that shift now are positioning themselves for the regulatory standard that is actively forming, rather than the one that is quietly becoming outdated.

About OmniTek & Contact Information

OmniTek Consulting is a trusted partner for organizations seeking to enhance governance, operational agility, and decision-making. We specialize in modernizing business processes with a blend of human-centered design, technology expertise, and hands-on change management.

For quality and data science leaders navigating AI model validation under an evolving regulatory framework, OmniTek applies proven best practices to build a practical credibility assessment practice: explicit context-of-use

documentation, risk-based validation tiering, and lifecycle monitoring built in from the start. Whether you are validating your first AI model or auditing an existing portfolio against the FDA and EMA's joint principles, OmniTek can help you build a practice that holds up under real regulatory scrutiny.

Contact us today to explore how OmniTek can support your transformation journey and drive lasting results.

McLean, VA | Serving clients nationwide
info@omnitekconsulting.com
omnitekconsulting.com

Click here to book a 15-minute call to see how OmniTek can help your team put these insights into practice.

Endnotes

[i] U.S. Food and Drug Administration, "FDA Proposes Framework to Advance Credibility of AI Models Used to Support Regulatory Decision-Making for Drug and Biological Products", available at https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions

[ii] IntuitionLabs, "FDA and EMA Guide to Good AI Practice in Drug Development," summarizing the January 2026 joint guiding principles and industry adoption data, available at https://intuitionlabs.ai/articles/fda-ai-drug-development-guidance