Journal of the Society for Clinical Data Management

Wartenberger M. Are You Ready for Elsa? What the FDA's Generative AI Inspection Tool Means for Clinical Data Teams. *Journal of the Society for Clinical Data Management*. 2026; 6(1): 4, pp. 1–5. DOI: https://doi.org/10.47912/jscdm.503

OPINION PAPER

Are You Ready for Elsa? What the FDA's Generative AI Inspection Tool Means for Clinical Data Teams

Marc Wartenberger

In June 2025, the U.S. Food and Drug Administration (FDA) launched its newest tool: Elsa is the FDA's internally deployed, large language model–based AI tool used to support regulatory review, data analysis, and operational efficiency. It is not formally defined as an acronym in FDA materials, but rather a named AI assistant integrated into agency workflows. This AI tool is designed to assist FDA staff with clinical protocol reviews, adverse event summarization, and inspection target identification. For clinical data management teams, Elsa signals a shift toward AI-assisted regulatory scrutiny that rewards well-structured, consistently documented, and traceable data. This opinion paper examines what Elsa does and does not do, identifies implications for data management practices, and connects these developments to the broader governance landscape established by the joint FDA–European Medicines Agency (EMA) Guiding Principles of Good AI Practice in Drug Development (January 2026) and the International Society for Pharmaceutical Engineering (ISPE) Good Automated Manufacturing Practice (GAMP) 5 Second Edition framework. We propose practical steps clinical data teams can take in 2026 to strengthen inspection readiness and position their organizations for an increasingly AI-enabled regulatory environment.

Keywords: Artificial Intelligence; FDA Elsa; Clinical Data Management; Inspection Readiness; Regulatory Compliance; Human Oversight; GAMP 5

1. Introduction

On June 2, 2025, the FDA officially launched Elsa, a generative AI tool built on a large language model (LLM) platform. Elsa is deployed agency-wide to assist FDA employees, from scientific reviewers to field investigators, in working more efficiently.[1]

The FDA described the rollout as the beginning of the agency's AI era, noting that tasks that previously took days could now be completed in minutes.[2] Elsa is already being used to expedite clinical protocol reviews, summarize adverse events, perform label comparisons, and identify high-priority inspection targets.[1,2,3]

Most clinical data management (CDM) teams, however, are not optimizing their data infrastructure with AI-assisted workflows in mind. The question is no longer whether clinical trial data will face AI-assisted review; it is whether that data will hold up under it. CDM organizations that invest now in documentation hygiene, audit trail integrity, and consistent metadata practices will be better positioned when regulators arrive, whether human, AI-assisted, or both.

This paper examines what Elsa does and does not do, identifies what these capabilities mean for clinical data teams specifically, connects these developments to the emerging regulatory governance framework, and proposes actionable steps for 2026 and beyond.

VP Compliance & Security, CRIO, US
marc.wartenberger@clinicalresearch.io

2. The Elsa Reality

Elsa is a LLM-powered AI tool operating within Amazon Web Services' secure GovCloud environment.[3] The FDA has emphasized that Elsa does not train on confidential data submitted by regulated industry; it queries only internal FDA data stores, which include adverse event databases such as the FDA Adverse Event Reporting System (FAERS) and Manufacturer and User Facility Device Experience (MAUDE), inspection histories, and clinical study records.[1,3,4]

At launch, the tool's capabilities included accelerating clinical protocol reviews and shortening the time needed for scientific evaluations, summarizing adverse events to support safety profile assessments, identifying high-priority inspection targets by analyzing patterns across regulatory data, performing faster drug label comparisons, and generating code for internal database development.[1,2,5] In December 2025, the FDA expanded its AI strategy further with an agency-wide agentic AI platform designed to support more complex, multi-step regulatory workflows while maintaining human oversight.[6]

It is important to understand Elsa's limitations. The tool has experienced accuracy challenges, including generating references to studies that do not exist, a well-documented hallucination risk inherent to large language models.[3,7] Anonymous FDA staffers have reported that Elsa cannot yet reliably assist with substantive review work, and external observers, including the law firm Hogan Lovells, have raised questions about what oversight governs the tool's outputs and whether benchmarks exist for evaluating its performance.[3,8] These limitations mean that Elsa's outputs function as investigative signals for FDA reviewers, not as final regulatory determinations. However, those signals will increasingly shape where and how deeply human reviewers focus their attention.

3. What this Means for Clinical Data Teams

3.1. Documentation Hygiene as a Strategic Priority

Elsa scans patterns across adverse events, protocol deviations, and regulatory submissions.[9] This has a direct implication: consistent terminology, complete metadata, and intact audit trails matter more than ever. When an AI tool parses thousands of adverse event records looking for safety signals, inconsistencies in coding conventions, free-text descriptions, or MedDRA term usage will surface as anomalies, regardless of whether they reflect genuine safety concerns or simply poor documentation practices.

As one industry advisor noted, sponsors, contract research organizations, and research sites should expect increased scrutiny of documentation quality and a greater emphasis on clarity, consistency, and data traceability across all submission materials.[10] For CDM teams, this means that documentation hygiene is no longer merely a quality metric; it is a competitive and compliance advantage. Organizations whose data is clean, traceable, and consistently structured will surface well under AI-assisted review. Those with fragmented, poorly documented, or inconsistently coded data will generate flags that invite deeper human investigation.

3.2. The Human-in-the-Loop Imperative

The FDA has stated that no enforcement action will be based solely on AI analysis without human review.[5] This principle of human oversight is not unique to the FDA's internal tools; it is a foundational expectation that extends to how regulated organizations use AI in their own operations. The joint FDA–EMA Guiding Principles of Good AI Practice in Drug Development, released on January 14, 2026, place human-centric design as the first of ten principles, emphasizing that AI systems must be built and deployed in ways that prioritize patient interests and include appropriate protections against foreseeable harm.[11,12,13]

For clinical data teams already using or evaluating AI tools, whether for data cleaning, coding automation, signal detection, or document generation, the question is straightforward: do your standard operating procedures explicitly define where human review is required before AI-generated outputs are acted upon? Ghosh and Huang (2025) frame this clearly in their discussion of LLMs in clinical trial data administration: AI tools should complement, not replace, human expertise, and their roles should be limited to providing support for routine tasks, with all critical decisions remaining the responsibility of qualified professionals.[13] Organizations should establish standard operating procedures (SOPs) that define acceptable AI applications, validation protocols, and boundaries.[14]

The risk analysis framework proposed by Ghosh and Huang is instructive here. They recommend assessing the nature of each risk when using LLMs for different tasks, evaluating the potential consequences if those risks materialize, and identifying the most sensitive prompts within clinical study workflows.[14] This three-part assessment, risk characterization, impact evaluation, and sensitivity mapping, provides a practical structure for CDM teams to evaluate their own AI-assisted processes.

3.3. Connecting Elsa to the Broader Regulatory Framework

Elsa does not exist in isolation. It is one expression of a broader regulatory shift toward AI-enabled oversight, and CDM teams should understand the governance landscape taking shape around it.

The FDA–EMA Guiding Principles, while currently non-binding, represent the shared regulatory thinking of both agencies and are explicitly intended to form the basis for future binding guidance.[11,12] The ten principles address ethical and human-centric design, risk-based approaches, robust data governance, multidisciplinary expertise, lifecycle management, and transparent communication about AI systems.[11,12,13] As one legal analysis observed, aligning early with these principles can reduce rework in submissions, support inspection readiness, and build confidence that AI-supported processes produce reliable, auditable results.[15]

Complementing this regulatory guidance, the ISPE GAMP 5 Second Edition (2022) and the companion ISPE GAMP Guide on Artificial Intelligence (July 2025) provide the operational validation framework.[9,16] GAMP 5 Second Edition introduced Appendix D11 on artificial intelligence/machine learning (AI/ML), which covered model development lifecycle, training data governance, performance metrics, and continuous monitoring.[9,16] The subsequent standalone AI guide extended these principles into a comprehensive 290-page framework for AI-enabled computerized systems in GxP environments.[16] Together with the FDA's finalized Computer Software Assurance guidance (September 2025), these frameworks provide CDM teams with a clear pathway for validating the AI tools they use internally.[16] See **Table 1** for a summary of key regulatory and governance frameworks.

CDM teams that have mapped their internal AI tools to the converging frameworks referenced in **Table 1** are meaningfully ahead of those who have not. This alignment is not merely about compliance; it is about demonstrating to regulators, sponsors, and auditors that AI-assisted processes are controlled, documented, and subject to appropriate human oversight.

Table 1: Key Regulatory and Governance Frameworks for AI in Clinical Data Management.

Framework	Date/Status	Relevance to CDM Teams
FDA–EMA Guiding Principles of Good AI Practice	January 2026; non-binding	Establishes 10 principles including human-centric design, risk-based approaches, data governance, and lifecycle management for AI used in drug development. Expected basis for future binding guidance.
ISPE GAMP 5 Second Edition	July 2022; current	Risk-based validation framework with AI/ML appendix (D11). Emphasizes Computer Software Assurance over legacy Computer Systems Validation (CSV). Foundation for validating AI tools used in CDM workflows.
ISPE GAMP Guide: Artificial Intelligence	July 2025; current	Comprehensive 290-page companion to GAMP 5 covering AI lifecycle from ideation through operation, including data governance, model risk management, and change control.
FDA Computer Software Assurance Guidance	September 2025; final	Endorses risk-based approach to software validation aligned with GAMP philosophy. Moves away from exhaustive documentation toward critical thinking and fit-for-purpose evidence.

4. Practical Steps for 2026 and Beyond

Based on the developments outlined above, we recommend the following actions for CDM organizations in 2026 and beyond:

1. **Audit your data for the inconsistencies Elsa is designed to surface**

Conduct an internal review of adverse event coding consistency, protocol deviation documentation, and metadata completeness across active studies. Identify patterns that an AI summarization tool would flag as anomalies and remediate them proactively.

Consider a multi-site oncology trial where Site A documents neutropenia-related dose modifications as protocol deviations using the term "dose reduction due to hematologic toxicity," Site B records the same clinical scenario as "dose held – ANC below threshold," and Site C logs it as a free-text narrative referencing the lab value without categorizing it as a deviation at all. Each site is documenting the same clinical event, but the inconsistent terminology, categorization, and completeness across the three records means that an AI tool scanning deviation logs and adverse event narratives in aggregate would identify these as potentially different safety patterns rather than a single, well-managed clinical scenario. The result is a flag, not because something went wrong clinically, but because the documentation does not tell a coherent story across sites.

This kind of terminology drift is common in multi-site and multi-regional trials, particularly where sites are working from different institutional templates or where CRF completion guidelines lack specificity on preferred terminology. A proactive audit should examine adverse event preferred term consistency across sites and coding dictionaries; deviation categorization alignment (ensuring comparable clinical events are classified using harmonized terminology); free-text narrative fields for undercoded or inconsistently described events; and metadata completeness in audit trails, particularly timestamps, user attribution, and reason-for-change entries. **Table 2** provides a practical audit checklist mapping common CDM data elements to the types of inconsistencies that AI summarization tools are designed to detect.

2. **Review SOPs to ensure human oversight of AI-assisted processes is explicitly documented**

Every SOP governing a process that incorporates AI, whether for data cleaning, coding, signal detection, or report generation, should specify the point at which a qualified professional reviews the output before it is acted upon. As Ghosh and Huang recommend, organizations should develop scoring guidelines from test prompts to categorize AI responses and clearly specify which tasks require human intervention.[14]

3. **Map current AI tools against the FDA–EMA risk framework**

For each AI tool in use or under evaluation, document its intended context of use, the risk classification based on the ten FDA–EMA principles, the validation approach aligned with GAMP 5 and the ISPE AI Guide, and the human oversight controls in place. This mapping exercise creates an auditable record that demonstrates governance maturity.

4. **Engage sponsor and CRO counterparts on shared inspection readiness**

Elsa's ability to scan across submissions means that inconsistencies between a sponsor's data and a CRO's documentation will surface together. Proactive alignment meetings with counterparts to harmonize coding conventions, audit trail standards, and deviation documentation practices can reduce inspection risk for all parties.

5. **Invest in cross-functional AI literacy**

Building internal understanding of how AI-assisted review works, and which signals it may elevate, benefits not only

Table 2: CDM Data Audit Checklist (Common Inconsistencies Surfaced by AI Summarization Tools).

Data Element	Common Inconsistency	Potential AI Flag	Recommended Audit Action
Adverse Event preferred terms	Different MedDRA terms used across sites for the same clinical event (e.g., "neutropenia" vs. "neutrophil count decreased")	Apparent divergence in safety profiles between sites that is actually a coding artifact	Review AE coding consistency across sites; confirm lowest-level term to preferred term mapping aligns with the current MedDRA version and study-specific coding conventions
Protocol deviation categorization	Same clinical scenario classified differently (e.g., one site logs a dose hold as a deviation, another as a clinical decision documented in the medical record only)	Uneven deviation rates across sites, suggesting either under-reporting at some sites or over-reporting at others	Harmonize deviation definitions in the CRF completion guidelines; provide site-specific examples in training materials
Free-text narrative fields	Inconsistent depth and terminology in verbatim descriptions of the same event type	Clusters of ambiguous or uncodeable narratives that do not align with structured field entries	Audit a sample of free-text entries against their corresponding coded fields; flag discrepancies for data management review
Audit trail metadata	Missing or incomplete reason-for-change entries, inconsistent timestamp formats, or user attribution gaps	Gaps in data provenance that undermine traceability, particularly for corrected or amended values	Run a system-level audit trail completeness report; verify that every data change carries a timestamp, user ID, and reason for change
Concomitant medication coding	Variations in drug name conventions (e.g. brand vs. generic vs. abbreviation) across sites or regions	Artificially inflated medication counts or missed drug interaction signals due to the same medication appearing as multiple distinct entries	Standardize medication coding to WHO Drug Dictionary preferred terms; reconcile regional brand name variations
Lab data units and reference ranges	Different units or reference ranges across central and local labs without clear normalization	Apparent out-of-range values that are actually unit conversion artifacts, or missed signals where values appear normal under one range but abnormal under another	Confirm unit standardization rules in the data management plan; validate that all local lab data has been normalized before analysis
Informed consent documentation	Inconsistent version tracking or date discrepancies between consent form versions and protocol amendments	Consent version mismatches that suggest participants may have been consented under outdated documents	Cross-reference consent version logs against IRB/EC approval dates and protocol amendment timelines

CDM teams but also regulatory, clinical operations, and quality assurance functions.[10] Cross-functional teams that understand the interaction between their data, AI review tools, and regulatory expectations can identify and resolve issues earlier in the submission lifecycle.

5. Conclusion

Elsa is the most prominent AI tool announcement by a regulatory body, but certainly not the last one. The FDA has already expanded beyond Elsa to agentic AI platforms capable of multi-step regulatory workflows,[6] and the EMA is advancing its own AI strategy aligned with the joint guiding principles.[11] The regulatory AI landscape will continue to evolve, and it would be prudent for CDM teams to assume that the scrutiny enabled by tools like Elsa will deepen, not diminish.

Organizations that build inspection-ready data infrastructure that is characterized by consistent documentation, explicit human oversight protocols, and validated AI governance frameworks will be better positioned as regulatory AI capabilities mature. The convergence of the FDA–EMA principles, GAMP 5, and the emergence of tools like Elsa creates a clear signal: the era of AI-assisted regulatory oversight has arrived, and clinical data teams should prepare accordingly.

Disclosures

The author acknowledges the use of Claude, an AI assistant developed by Anthropic, to support research, source identification, and drafting during the preparation of this manuscript. All content was subsequently reviewed, verified for factual accuracy, and edited by the author to ensure alignment with professional standards and the author's own analysis and opinions. The use of AI-assisted tools in the preparation of this work is consistent with the principles of human oversight discussed within the paper itself.

Competing Interests

The author has no competing interests to declare.

References

1. **Kelley A.** FDA unveils 'Elsa' generative AI tool for staff. Nextgov/FCW; 2025, June 3. https://www.nextgov.com/artificial-intelligence/2025/06/fda-unveils-elsa-generative-ai-tool-staff/405761/ Accessed January 15, 2026.
2. **U.S. Food and Drug Administration.** FDA Launches Agency-Wide AI Tool to Optimize Performance for the American People. FDA News Release; 2025, June 2. https://www.fda.gov/news-events/press-announcements/fda-launches-agency-wide-ai-tool-optimize-performance-american-people Accessed January 15, 2026.
3. **Applied Clinical Trials.** FDA's Elsa AI Tool Raises Accuracy and Oversight Concerns. Applied Clinical Trials Online; 2025. https://www.appliedclinicaltrialsonline.com/view/fda-elsa-ai-tool-raises-accuracy-and-oversight-concerns Accessed April 1, 2026.
4. **Witowski N.** How AI tool Elsa could shape the FDA review process. Definitive Healthcare Blog; 2025, July 15. https://www.definitivehc.com/blog/fda-releases-ai-tool-elsa Accessed January 15, 2026.
5. **Berman JR, Seely AZ.** FDA's AI-Assisted Review: The Next Stage of Regulated Product Evaluation. As Prescribed Blog; 2025, June 3. https://www.morganlewis.com/blogs/asprescribed/2025/06/fdas-ai-assisted-review-the-next-stage-of-regulated-product-evaluation Accessed January 15, 2026.
6. **Studna A.** FDA Deploys Secure Agentic AI Platform to Modernize Regulatory Operations. Applied Clinical Trials Online; 2025. https://www.appliedclinicaltrialsonline.com/view/fda-deploys-secure-agentic-ai-platform-modernize-regulatory-operations Accessed April 1, 2026.
7. **Masters K.** Medical Teacher's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers. *Medical Teacher*. 2023;45(7):673–675. DOI: https://doi.org/10.1080/0142159X.2023.2208731
8. **Hogan Lovells.** FDA launches "Elsa" AI tool to aid drug approvals. Hogan Lovells Publications; 2025, June 12. https://www.hoganlovells.com/en/publications/fda-launches-elsa-ai-tool-to-aid-drug-approvals Accessed January 15, 2026.
9. **ISPE.** GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems (Second Edition). International Society for Pharmaceutical Engineering; 2022. https://ispe.org/publications/guidance-documents/gamp-5-guide-2nd-edition Accessed April 1, 2026.
10. **Advarra.** Preparing for ELSA: What Sponsors, Sites, and CROs Need to Know About the FDA's New AI Era. Advarra Blog; 2026, January 28. https://www.advarra.com/blog/preparing-for-elsa-what-sponsors-sites-cros-need-to-know-about-fdas-new-ai-era/ Accessed April 1, 2026.
11. **European Medicines Agency.** EMA and FDA set common principles for AI in medicine development. EMA News; 2026, January 14. https://www.ema.europa.eu/en/news/ema-fda-set-common-principles-ai-medicine-development-0 Accessed April 1, 2026.
12. **McGuireWoods.** FDA and EMA Provide Guiding Principles for AI in Drug Development. McGuireWoods Client Resources; 2026, January 27. https://www.mcguirewoods.com/client-resources/alerts/2026/1/fda-and-ema-provide-guiding-principles-for-ai-in-drug-development/ Accessed April 1, 2026.
13. **ArentFox Schiff.** Model Behavior: FDA and EMA's Guide to Good AI in Drug Development. AFS Perspectives; 2026, February 9. https://www.afslaw.com/perspectives/alerts/model-behavior-fda-and-emas-guide-good-ai-drug-development Accessed April 1, 2026.
14. **Ghosh A, Huang CJ.** Leveraging Large Language Models to Streamline Clinical Trial Data Administration: Balancing Efficiency with Ethical Responsibility. *Journal of the Society for Clinical Data Management*. 2025;5(1):18, pp. 1–8. DOI: https://doi.org/10.47912/jscdm.438
15. **Jones Day.** EMA and FDA Align on Good AI Practice in Drug Development. Jones Day Insights; 2026, January 28. https://www.jonesday.com/en/insights/2026/01/ema-and-fda-align-on-good-ai-practice-in-drug-development Accessed April 1, 2026.
16. **IntuitionLabs.** GAMP 5 Second Edition: A Guide to Key Changes & Updates. IntuitionLabs; 2026. https://intuitionlabs.ai/articles/gamp-5-second-edition-updates Accessed April 1, 2026.

How to cite this article: Wartenberger M. Are You Ready for Elsa? What the FDA's Generative AI Inspection Tool Means for Clinical Data Teams. *Journal of the Society for Clinical Data Management*. 2026; 6(1): 4, pp. 1–5. DOI: https://doi.org/10.47912/jscdm.503

Submitted: 18 April 2026 **Accepted:** 04 May 2026 **Published:** 25 May 2026

JSCDM *Journal of the Society for Clinical Data Management* is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS