Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

FDA and EMA on AI in Medicines: Credibility and Traceability

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · fda ai guidance drug development · ema reflection paper ai · ai credibility framework fda · context of use ai model fda · ai in pharmacovigilance regulation

Original article: https://reliablecontext.com/articles/fda-ema-ai-credibility-regulatory-decision-making-medicines



In this article

1. Two agencies, one question
2. Terms used in this article
3. FDA's draft guidance: the credibility assessment framework
4. EMA's reflection paper: a lifecycle view
5. The joint principles of January 2026
6. Comparing the three
7. A worked example: what traceability means for one output
8. Interaction with Part 11, GCP and existing data integrity rules
9. What remains unsettled
10. Summary
11. Sources

Two agencies, one question

When a sponsor uses an AI model to produce evidence about a medicine, the regulator has to decide whether it can rely on that evidence. The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have approached the question through different instruments, but their answers converge on three ideas: define precisely what the model is being used for, scale the scrutiny to the consequences of an error, and keep a traceable record of the data and the model that produced each output.

This article reads three documents side by side:

Document	Issuer	Status on 28 October 2026
Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products	FDA (CDER, CBER, CDRH, CVM, OCE, OCP and others)	Draft guidance, docket FDA-2024-D-4689; the FDA page still labels it "Draft"
Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle, EMA/CHMP/CVMP/83833/2023	EMA (CHMP and CVMP)	Final; adopted by CHMP on 9 September 2024
Guiding principles of good AI practice in drug development	FDA and EMA jointly	Published January 2026; ten high-level principles

None of these documents is a regulation. FDA states that its guidance documents "do not establish legally enforceable responsibilities" (FDA draft guidance, Introduction). This article is educational and is not regulatory or legal advice.

Terms used in this article

- **Context of use (COU).** FDA: "The COU defines the specific role and scope of the AI model used to address a question of interest" (FDA draft guidance). Role is what the model does; scope is how much the decision depends on it.

- **Question of interest.** The "specific question, decision, or concern being addressed by the AI model" (FDA, Step 1).
- **Credibility.** "Trust, established through the collection of credibility evidence, in the performance of an AI model for a particular COU" (FDA, Introduction).
- **Model risk.** In the FDA framework, a combination of *model influence* (the contribution of the model's evidence relative to other evidence) and *decision consequence* (the significance of an adverse outcome resulting from an incorrect decision) (FDA, Step 3).
- **High patient risk / high regulatory impact.** EMA's two labels: the first for "systems affecting patient safety", the second for cases where "impact on regulatory decision-making is substantial" (EMA reflection paper, 2.2).

FDA's draft guidance: the credibility assessment framework

Scope

The draft covers AI models used "to produce information or data to support regulatory decision-making regarding safety, effectiveness, or quality for drugs". It explicitly does not address AI used in drug discovery, or AI "used for operational efficiencies (e.g., internal workflows, resource allocation, drafting/writing a regulatory submission)" that does not affect patient safety, drug quality or the reliability of study results (FDA draft guidance, section II). A model that drafts a clinical study report is therefore out of scope; a model whose output decides which trial participants go home after dosing is in scope.

The seven steps

Step	FDA wording (abbreviated)	Output a sponsor would hold
1	Define the question of interest	One sentence naming the decision
2	Define the COU	Role, scope, and other evidence used alongside the model
3	Assess the AI model risk	Model influence and decision consequence ratings, with reasoning
4	Develop a plan to establish credibility within the COU	Credibility assessment plan
5	Execute the plan	Results
6	Document results and discuss deviations	Credibility assessment report
7	Determine the adequacy of the model for the COU	Decision, or one of the remediation options

Source: FDA draft guidance, section IV.A.

The guidance gives two hypothetical examples. In the clinical one, a model stratifies participants for 24-hour inpatient monitoring after dosing a drug with a life-threatening adverse reaction, and "only the AI model will be used to determine whether the participant is considered low risk"; FDA rates both model influence and decision consequence high. In the manufacturing one, a model assesses vial fill volume, but independent verification is

performed on a representative sample of each batch, so the model "will not be the sole determinant for the release of product" and model influence is rated low. The contrast is instructive: the same kind of model, under a different COU, carries a different risk.

Data expectations inside the plan

Step 4 is where data enters. FDA asks that the data used to develop the model be "fit for use", meaning "both relevant ... and reliable (i.e., accurate, complete, and traceable)" (FDA draft guidance, IV.A.4). Commensurate with model risk, the plan should describe how development data were "collected, processed, annotated, stored, controlled, and used", how labels or annotations were established, how data were split into training, tuning and other subsets, and which development activities used which data set. For test data, it asks the sponsor to specify "how data independence was achieved between development ... and test data", and to justify any overlap. It also asks for "the quality assurance and control procedures of computer software ... and how version changes were tracked".

FDA's footnote on terminology is worth noting: some in the AI community call tuning data "validation" data, but "FDA does not use the word validation in this context".

Steps 6 and 7

The results go into a credibility assessment report, which may be submitted or "held and made available to FDA on request (e.g., during an inspection)". If credibility is not sufficiently established, Step 7 lists five outcomes: reduce model influence by adding other evidence, increase the rigour of the assessment or add development data, add controls, change the modelling approach, or treat the model's output as inadequate for the COU (FDA draft guidance, IV.A.7).

Life cycle maintenance

FDA singles out models whose use extends over time, such as in manufacturing. "Model performance metrics should be monitored on an ongoing basis", and the level of oversight "should be risk-based (i.e., commensurate with the model risk and the COU)" (FDA draft guidance, IV.B). Changes that may affect performance are expected to go through the manufacturer's change management system within its pharmaceutical quality system.

EMA's reflection paper: a lifecycle view

The EMA paper is organised by phase of the medicinal product lifecycle: drug discovery, non-clinical development, clinical trials, precision medicine, product information, manufacturing and the post-authorisation phase (EMA reflection paper, table of contents). Its central allocation of responsibility is stated plainly: it is "the responsibility of the clinical trial sponsor, marketing authorisation applicant/holder or manufacturer to ensure that all algorithms, models, datasets, and data processing pipelines used are fit for purpose", and these requirements "may in some respects be stricter than what is considered standard practice in the field of data science" (2.2).

Four provisions matter most for data and traceability:

1. **Clinical trial records.** Where a use "could be of high regulatory impact or high patient risk in a clinical trial" and the method has not been qualified by EMA for that COU, "the full model architecture, logs from model development, validation and testing, training data and description of the data processing pipeline would likely be considered parts of the clinical trial data or trial protocol dossier" and may be requested at marketing authorisation, clinical trial application or GCP inspection (2.3.3.1).
2. **Frozen pipelines in pivotal trials.** Before database lock and unblinding, "the data pre-processing pipeline and all models should be frozen and documented in a traceable manner in the statistical analysis plan". Non-prespecified changes after the data set is opened make results post hoc. "Incremental learning approaches are not accepted" in that setting (2.3.3.3).
3. **Data acquisition.** "The source(s) of data and the process of data acquisition, along with any processing such as cleaning, transformation, imputation, annotation, normalisation, and augmentation should be documented in a detailed and fully traceable manner in line with GxP requirements" (2.5.1).
4. **Governance.** "SOPs implementing GxP principles on data and algorithm governance should be extended to include all data, models and algorithms used for AI/ML in cases of high regulatory impact or high patient risk" (2.6).

Pharmacovigilance is treated differently. The paper allows "a more flexible approach to AI/ML modelling and deployment, where incremental learning can continuously enhance models" for adverse event classification and signal detection, while it "remains the responsibility of the MAH to validate, monitor and document model performance" (2.3.7). On large language models specifically, the paper's comment is about data protection: they "are at particular risk of memorisation due to their size" (2.7).

The joint principles of January 2026

In January 2026 the two agencies published ten guiding principles. Principle 4, "Clear context of use", and Principle 2, "Risk-based approach ... based on the context of use and determined model risk", restate the FDA vocabulary. Principle 6, "Data governance and documentation", is the one closest to traceability: "Data source provenance, processing steps, and analytical decisions are documented in a detailed, traceable, and verifiable manner, in line with GxP requirements" (FDA and EMA, Guiding principles). Principle 8 asks that performance assessments evaluate "the complete system including human-AI interactions", and Principle 9 asks for "scheduled monitoring and periodic re-evaluation ... (e.g., to address data drift)".

Comparing the three

Topic	FDA draft guidance	EMA reflection paper	Joint principles
Risk concept	Model influence x decision consequence	High patient risk, high regulatory impact	Risk-based, by COU and model risk
Unit of assessment	Model output for a COU	Algorithms, models, data sets and pipelines	AI technology across the life cycle
Data quality words	Relevant; reliable = accurate, complete, traceable	Fully traceable in line with GxP	Detailed, traceable, verifiable

Topic	FDA draft guidance	EMA reflection paper	Joint principles
Change after deployment	Life cycle maintenance, change management	Frozen in pivotal trials; incremental learning allowed in pharmacovigilance	Scheduled monitoring and re-evaluation
Human in the loop	Evaluate the human-AI team	Explainability methods whenever possible	Evaluate the complete system including human-AI interactions
Documentation	Credibility assessment plan and report	Model architecture, development logs, training data, pipeline description	Documentation of provenance, processing steps, analytical decisions

A worked example: what traceability means for one output

The example is fictional. Halden Therapeutics AG (a fictional company) uses a model to flag likely serious adverse events among incoming case reports so that medical reviewers see them first. On 12 May 2026 the model scores case PV-2026-0381 at 0.91 and the case is escalated.

Applying the FDA steps: the question of interest is "Which incoming reports should be reviewed first for seriousness?" The COU is triage only; every case is still reviewed by a person, so model influence is lower than if the model decided seriousness on its own. The decision consequence of a missed serious case is high, because reporting timelines depend on it.

Six months later an inspector asks why a similar case, PV-2026-0402, was not escalated. To answer, Halden needs to reconstruct the conditions of that one output:

Element	What must be retrievable	Why
Input	The exact case text the model received, after any de-identification or translation step	FDA asks how data were processed; EMA asks processing to be fully traceable
Model version	Identifier of the deployed model and its training data set version	EMA 2.5.6 expects re-evaluation after non-trivial changes
Pipeline version	Pre-processing code version, including tokenisation and field mapping	EMA 2.5.1 lists transformation and normalisation
Output	Score, threshold in force that day, and the decision	FDA Step 7 requires a defined adequacy decision
Human action	Reviewer, time, and whether they overrode the triage	Joint principle 8: the human-AI system is the unit
Monitoring	Performance metrics for the period, and any drift alerts	FDA IV.B; joint principle 9

If the input text was not retained, or the threshold changed without a record, the sponsor can describe its system in general but cannot explain this output. Measuring such a model against a labelled reference set, as described in the site's method note on evaluating extraction against a gold set, only helps if each scored item can be linked back to the version that scored it.

Interaction with Part 11, GCP and existing data integrity rules

These AI documents sit on top of existing rules rather than replacing them. EMA says AI used in clinical trials "should meet applicable requirements in the ICH E6 guideline for good clinical practice" (2.3.3.1), and ICH E6(R3) section 4.3.4 requires that protocol-specific configurations "including automated data entry checks and calculations" be validated (ICH E6(R3)). In the United States, a computerised system that creates or modifies electronic records within Part 11 scope must use "secure, computer-generated, time-stamped audit trails" (21 CFR 11.10(e)). A model's outputs that are written into such records inherit those obligations; the model's own training logs are a separate matter that the AI documents address.

The pharma sector note on this site describes the general requirement behind all of this: each value a regulated decision rests on needs a path back to its source.

What remains unsettled

- **Final FDA text.** The FDA guidance was still a draft on 28 October 2026. Specific recommendations, such as the requirement that "all performance estimates should be provided with confidence intervals", may change.
- **Generative models.** The FDA draft and the EMA paper were written mainly with predictive models in mind. Neither sets out how to define a COU, or what "test data independence" means, for a general-purpose language model used in a regulated workflow.
- **Device boundary.** FDA notes that an AI model may meet the definition of a device depending on intended use, and that devices are addressed by device-specific guidance; EMA refers medical devices to the device regulations.

Summary

FDA asks sponsors to define the question, the context of use and the model risk, and then to plan, execute and document a credibility assessment proportionate to that risk. EMA asks the same parties to own every algorithm, data set and pipeline, to freeze them where confirmatory evidence is at stake, and to document data acquisition and processing in a fully traceable way. The 2026 joint principles state the common core: provenance, processing steps and analytical decisions documented in a traceable and verifiable manner.

Sources

1. U.S. Food and Drug Administration, *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products*, draft guidance, docket FDA-2024-D-4689 (web page). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
2. U.S. Food and Drug Administration, same draft guidance, PDF. https://www.fda.gov/media/184830/download

3. European Medicines Agency, *Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle*, EMA/CHMP/CVMP/83833/2023, 9 September 2024. https://www.ema.europa.eu/system/files/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle-en.pdf
4. European Medicines Agency and U.S. Food and Drug Administration, *Guiding principles of good AI practice in drug development*, January 2026. https://www.ema.europa.eu/en/documents/other/guiding-principles-good-ai-practice-drug-development_en.pdf
5. ICH, *Guideline for Good Clinical Practice E6(R3)*, 6 January 2025. https://database.ich.org/sites/default/files/ICH_E6(R3)_Step4_FinalGuideline_2025_0106.pdf
6. Electronic Code of Federal Regulations, 21 CFR 11.10, Controls for closed systems. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11/subpart-B/section-11.10

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
2. https://www.ema.europa.eu/system/files/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle-en.pdf
3. https://www.ema.europa.eu/en/documents/other/guiding-principles-good-ai-practice-drug-development_en.pdf
4. https://www.fda.gov/media/184830/download
5. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
6. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11/subpart-B/section-11.10

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.