

FDA's AI Guidance: 7-Step Credibility Framework Explained

1/2/2026 • 50 min read

- fda guidance
- ai in drug development
- credibility framework
- regulatory affairs
- model validation
- context of use
- risk-based approach
- biologics regulation
- machine learning



[Revised April 21, 2026] This article has been updated to reflect developments since original publication, including the close of the FDA's public comment period (April 7, 2025), the January 2026 joint FDA–EMA "Guiding Principles of Good AI Practice in Drug Development," and the FDA's expected Q2 2026 finalization of the draft guidance.

Executive Summary

In January 2025 the U.S. Food and Drug Administration (FDA) issued a draft guidance on the use of artificial intelligence (AI) in drug and biological product development, establishing a comprehensive **7-step credibility assessment framework** (^[1] www.fda.gov) (^[2] pmc.ncbi.nlm.nih.gov). This framework mandates that sponsors clearly define the **regulatory question** the AI model is intended to address and its **context of use (COU)**, assess model **risk** (based on *model influence* and *decision consequence*), develop and execute a **credibility assessment plan** aligned with that risk, and document the outcomes. At its core, the approach is *risk-based* – higher-risk AI applications (e.g., those directly impacting patient safety or product quality) require more rigorous validation, extensive documentation, and possibly lifecycle monitoring, whereas lower-risk uses may need only basic evidence (^[3] www.ark-biotech.com) (^[4] www.wcgclinical.com).

Our report provides an in-depth analysis of this framework. We trace the historical context of AI in drug development and regulatory engagement, then unpack each of the seven steps with detailed explanation and examples. We draw on FDA source documents and expert analyses to illustrate how sponsors should define the question of interest, articulate the COU, quantify model risk, plan and carry out validation, and decide if the model's performance is adequate. Crucial themes include the need for **fit-for-use, high-quality data**, thorough performance metrics (e.g. accuracy, bias checks, uncertainty quantification), and **documentation** of any deviations or life-cycle changes (^[4] www.wcgclinical.com) (^[5] www.foley.com).

We also examine multiple perspectives: how the industry is preparing (including case examples from bioprocess simulation and **pharmaceuticals manufacturing**), concerns about intellectual property versus transparency (^[6] www.foley.com), and parallels with international guidelines (e.g. EMA's AI reflection paper) and academic recommendations for AI model validation (^[7] pmc.ncbi.nlm.nih.gov) (^[8] www.nature.com). For instance, an Ark Biotech case study applies the framework to a bioreactor simulation for antibody production, showing how an AI-driven "hybrid" model identified optimal process parameters and replaced physical trials (^[9] www.ark-biotech.com). Stakeholders from law firms to healthcare system CIOs emphasize early FDA engagement and robust risk management as best practices (^[10] www.dlapiper.com) (^[11] www.foley.com).

Finally, we discuss broader implications. The guidance's emphasis on transparency and documentation spurs innovation (e.g. explainable models, automated reporting tools) (^[5] www.foley.com) and may influence intellectual property strategies. Future directions include finalizing the guidance after public comments, global harmonization (the FDA is already working with other regulators), and ongoing FDA support initiatives. In sum, the FDA's framework outlines how to **build trustworthy AI in drug development**, requiring a clear rationale, strong evidence, and continuous oversight to ensure patient safety and product quality while fostering innovation (^[12] www.fda.gov) (^[13] pmc.ncbi.nlm.nih.gov).

Introduction and Background

Artificial intelligence is increasingly transforming biomedical research and drug development. Machine learning models and advanced algorithms are now used at every stage – from *in silico* compound screening and trial design to **manufacturing process optimization** and post-market surveillance (^[14] pmc.ncbi.nlm.nih.gov) (^[15] pmc.ncbi.nlm.nih.gov). The FDA has long encouraged innovation, reviewing over 300 submissions with AI components by 2023 (a figure that continued to climb sharply through 2025 as AI tooling matured) (^[15] pmc.ncbi.nlm.nih.gov) (^[12] www.fda.gov) and running workshops (e.g. with Duke's CTTI in 2024 (^[16] pmc.ncbi.nlm.nih.gov)) and public discussions on AI. Yet regulators and

sponsors have struggled with how to ensure these models are **reliable and safe** when making high-stakes decisions. Recognizing both the promise and perils of AI, the FDA in early 2025 released its first guidance specifically on **AI in drug and biologic development** (^[1] www.fda.gov) (^[17] www.futuremedicine.com).

The FDA's draft guidance – titled *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products* (FDA-2024-D-4689) – is part of this ongoing effort. It complements earlier AI-related guidance (e.g. for AI-enabled medical devices in 2023) by focusing on drugs/biologics. The document provides recommendations to sponsors on using AI to produce data or information that will support regulatory decisions about *safety, effectiveness or quality* (^[18] www.fda.gov). In doing so, it introduces a **risk-based credibility assessment framework** for AI models, signaling that trust in an AI model's output must be established *before* the FDA will consider it in a regulatory submission (^[19] regulations.justia.com) (^[2] pmc.ncbi.nlm.nih.gov).

Key motivations for this guidance include: (1) **Regulatory experience** – since 2016 the FDA has seen a sharp rise in AI applications in submissions, now totaling hundreds of cases (^[12] www.fda.gov) (^[15] pmc.ncbi.nlm.nih.gov); (2) **Stakeholder input** – the guidance builds on an FDA-led workshop (Duke Margolis, 2022) and over 800 public comments on prior AI discussion papers in 2023 (^[20] www.fda.gov) (^[11] www.foley.com); and (3) **Public health** – given the impact AI could have on patient safety and drug quality, FDA needs clear policies that balance innovation with its statutory safety standards (^[21] www.fda.gov). As Commissioner Califf explained, “With the appropriate safeguards in place, AI has transformative potential to advance medical product development”—but those safeguards must include robust technical and regulatory checklists (^[21] www.fda.gov).

Conceptual framing. In this guidance, FDA defines an AI system broadly as any “machine-based system that can... make predictions, recommendations, or decisions influencing real or virtual environments” (^[22] www.wcgclinical.com). The focus is on *maps of reality* – data-driven models that produce outputs (e.g. risk scores, image classifications) intended to influence a regulatory question. Importantly, the guidance repeatedly emphasizes **context of use (COU)**, a term FDA uses to describe “how and in what context” a model addresses a specific question (^[23] www.fda.gov) (^[19] regulations.justia.com). COU includes the model's scope, its inputs/outputs, and its role relative to other evidence. It also defines **credibility** in a regulatory sense: “trust, established through the collection of credibility evidence, in the performance of an AI model for a particular COU” (^[19] regulations.justia.com). Thus, an AI model is not inherently acceptable or not – its credibility depends on whether there is trustworthy evidence that it works for *the specified decision task*.

The guidance expressly excludes certain AI uses: it does *not* apply to AI for early drug discovery (e.g. target ID or lead optimization) or purely internal business tasks. Instead, it applies to situations where an AI-generated output will be used in a submission or quality system that directly affects patient safety, drug efficacy, or manufacturing/compliance decisions (^[23] www.fda.gov) (^[24] www.foley.com). Examples include AI-driven trial designs, predictive models for patient outcomes, R&D tools for digital biomarkers, pharmacovigilance analytics, and manufacturing process controls (^[25] www.dlapiper.com) (^[26] www.wcgclinical.com). But if AI is used only to speed engineering of a molecule *before* safety/efficacy are assessed, it lies outside the guidance's scope (^[24] www.foley.com).

In summary, the FDA guidance aims to **advance trustworthy AI** in drug development by requiring sponsors to plan and document how they will validate and monitor models in context. It essentially offers a roadmap – the seven steps – for building a regulatory submission with credible AI. The rest of this report explains each aspect in depth, showing how the framework works and what it means for the future of drug development.

The 7-Step Credibility Assessment Framework

The heart of the FDA's draft guidance is a **seven-step, risk-based framework** to evaluate the credibility of AI models for a defined COU (^[2] pmc.ncbi.nlm.nih.gov) (^[27] www.dlapiper.com). Below we explain each step, synthesizing regulatory text, expert commentary, and concrete examples. (A summary table of the steps is provided at the end of this section.)

Step 1: Define the Question of Interest

Before building or validating any AI model, developers must *clearly articulate the regulatory question* it is intended to address (^[23] www.fda.gov) (^[28] www.dlapiper.com). This step sets the stage for everything that follows. It involves specifying the precise decision or outcome the AI's output will influence, aligned with FDA-defined objectives such as safety, efficacy, or quality.

For example, in a clinical trial setting the question might be, "Which trial participants are sufficiently low-risk that they need only outpatient monitoring after receiving Drug X in study Y?" (^[29] www.bioprocessonline.com). That question implicates patient safety (are we at risk of missing adverse events?) and guides the model's design. In manufacturing, a question could be, "Do the vials produced in batch Z of Drug Y meet our fill-volume specifications?" (^[30] www.bioprocessonline.com). This concerns product quality and compliance.

Defining the question of interest forces sponsors to tie their AI use-case to a concrete decision. As Foley LLP notes, examples could include trial inclusion criteria, patient risk stratification, or adjudicating clinical endpoints (^[31] www.foley.com). At this stage, sponsors should also identify *contextual evidence* or background data supporting the question – for instance, historical trial results or prior process performance. Crucially, the question determines the **intended use**: whether the AI output will "inform" a decision (with human review) or actually *make* the decision autonomously. As the FDA highlights, models making final determinations without human oversight generally carry higher risk and thus demand stronger evidence (^[4] www.wcgclinical.com).

In summary, Step 1 ensures that the AI effort is problem-driven. By articulating the question of interest, sponsors define the model's **purpose and boundaries**. All subsequent planning – from defining inputs and performance criteria to assessing risk – centers on this question. Without a specific question, the framework cannot be applied.

Step 2: Define the Context of Use (COU)

Having identified the question, Step 2 requires sponsors to describe *how the AI model will be used to address it*, i.e. its **context of use** (^[32] www.dlapiper.com) (^[33] www.wcgclinical.com). COU includes the operational details and scope surrounding the model: what data it ingests, what output it produces, and how that output will be integrated into decision-making. It also specifies any boundaries (e.g. patient populations, manufacturing conditions) in which the model is expected to work.

Precisely defining the COU is critical because it informs the level of stringency needed in validation. For instance, if the COU specifies that the model's output will *only be one input among many* in a decision process (with a human ultimately accountable), the risk is lower than if the model's result will directly determine an outcome. As one analyst notes, the COU description should spell out whether other information (e.g. lab tests, human review) will complement the AI output, or if the AI is the sole decision-maker (^[34] www.dlapiper.com) (^[35] www.wcgclinical.com).

Concrete examples illustrate COU definitions. In a clinical scenario, the COU might be: "The AI model will assign participants to low- or high-risk groups for adverse events based on baseline data (age, labs, genomics), thereby informing monitoring intensity; human clinicians will then apply additional judgment (^[36] www.bioprocessonline.com)."
In a manufacturing context, the COU could be: "Analyze real-time process sensor data to detect deviations in product quality attributes (e.g. fill level); any out-of-spec predictions will be reviewed by quality staff before any corrective action (^[37] www.bioprocessonline.com)."

In this step, sponsors also outline **data requirements** for the COU. They must identify sources of training, testing, and validation data that are "fit for use" – i.e. relevant and reliable for the intended purpose (^[19] regulations.justia.com) (^[38] pmc.ncbi.nlm.nih.gov). Data quality criteria (completeness, consistency, etc.) and lifecycle data flows should be detailed. The plan should cover whether real-world data, historical trial data, or prospective data will be used, and how data

pivotal points like missingness or noise will be handled. The FDA emphasizes that clear data validation processes are needed, especially when using diverse data (such as RWD) (^[39] www.bioprocessonline.com).

Completing Step 2 yields a comprehensive COU statement that defines the *embedded role* of AI. This includes: **input types, output interpretation, integration with other evidence, operational procedures, environmental constraints, and user interactions**. This clarity is the foundation for risk analysis: once the COU is set, sponsors can judge what could go wrong when the model is used as intended.

Step 3: Assess Model Risk

Step 3 applies a **risk assessment** to the AI model within its COU (^[40] www.dlapiper.com) (^[41] www.wcgclinical.com). FDA instructs sponsors to evaluate two core vectors of risk:

- **Model Influence** – how strongly will the AI output influence the decision or process? Is it *one factor* among many, or does it solely determine a critical outcome? An AI that operates as a “black box” with full authority is highly influential (and thus higher risk) because its errors directly propagate to decisions (^[42] www.wcgclinical.com). By contrast, a model whose output is only advisory or is double-checked by humans has lower influence.
- **Decision Consequence** – what is the potential harm if the AI output is wrong? This measures severity of an incorrect decision in the given application. For a clinical model predicting adverse events, incorrect stratification could endanger patient safety (high consequence). In manufacturing, an erroneous prediction about a non-critical parameter might have lower consequence (though some failures could still affect product quality) (^[43] www.wcgclinical.com).

Combining these factors yields a model risk classification (e.g. Low, Medium, High). As DLA Piper explains, sponsors should consider both the *severity* of potential errors and the *probability* of occurrence, possibly using tools like risk matrices or Failure Modes and Effects Analysis (FMEA) (^[44] www.bioprocessonline.com) (^[45] www.dlapiper.com). For example, a scenario with **high influence & high consequence** (AI makes an unverified decision on patient monitoring, where an error could be life-threatening) is categorized as high-risk (^[46] www.bioprocessonline.com) (^[47] www.wcgclinical.com). By contrast, **low influence & low consequence** scenarios (AI assists manufacturing QC but humans ultimately verify) are low-risk. Intermediate combinations yield medium risk.

Assessing risk also involves considering non-clinical threats. Sponsors should evaluate data integrity (training/test splits, label quality), bias (e.g. non-representative samples), adversarial vulnerabilities (e.g. input manipulation), and cybersecurity risks (^[48] www.bioprocessonline.com). For instance, bias detection should be an integral part of risk analysis, with plans for fairness evaluation across subgroups. Algorithms like Bayesian models, which inherently estimate uncertainty, can inform the risk profile by quantifying confidence (^[49] www.bioprocessonline.com) (^[50] www.bioprocessonline.com).

Quantifying model risk has two uses. First, it informs the depth of subsequent credibility activities: *higher-risk models require more evidence*. The guidance explicitly ties risk to the stringency of validation and documentation (^[51] www.foley.com) (^[52] www.dlapiper.com). Second, it highlights where risk mitigation may be needed (e.g. adding human review, reducing model reliance). In fact, Step 7 (below) eventually checks adequacy of the risk assessment and whether mitigations or evidence supplements are required.

Step 4: Develop a Credibility Assessment Plan

With risk classified, Step 4 directs sponsors to **plan** how they will demonstrate the model's credibility for the COU (^[53] www.dlapiper.com) (^[54] www.wcgclinical.com). This plan is essentially a project blueprint for all validation and verification activities. It should be tailored to the inferred risk level: the higher the risk, the more comprehensive and rigorous the plan.

The plan must enumerate all major elements of the AI model and its development lifecycle: model description (type, architecture, assumptions), data descriptions, training/tuning procedures, and proposed evaluation methods. For example, sponsors should specify which data sets will be used for model development vs testing, how data quality will be assessed, and what metrics or acceptance criteria will be used to judge performance ⁽⁴⁶⁾ www.dlapiper.com ⁽⁴⁷⁾ www.wcgclinical.com). Performance metrics might include accuracy, sensitivity/specificity, predictive values, calibration statistics, and uncertainty measures; appropriate choices depend on the task (e.g. a binary safety prediction model vs a continuous process control output). The plan should also describe any strategies for oversampling, cross-validation, or external validation to ensure generalizability.

The draft guidance lists key plan components explicitly: description of the model and data, description of training/tuning data, description of test data, description of model evaluation process, etc ⁽⁴⁶⁾ www.dlapiper.com. Sponsors should address potential *sources of error or bias* in each component (for example, noting any limitations in training data representativeness and how those will be mitigated). Crucially, life-cycle planning should be included: if the model will learn or drift over time, the plan should outline how continuous monitoring and revalidation will occur ⁽³⁵⁾ www.wcgclinical.com ⁽⁴⁸⁾ www.foley.com).

Developing the plan is often iterative; sponsors are encouraged to engage FDA at this stage to align expectations. By planning before full implementation, sponsors can adjust their approach (e.g. gather more data, add controls) if the proposed activities seem insufficient. For high-risk models especially, the plan might call for pre-specified success criteria (e.g., thresholds for false-positive/negative rates) and independent review points.

Step 5: Execute the Credibility Plan

Step 5 is execution: sponsors carry out the planned validation and testing activities ⁽⁴⁹⁾ www.dlapiper.com. In practical terms, this means training the model on the development data, rigorously testing it on hold-out or external datasets, tuning parameters, and conducting any additional analyses such as sensitivity or stress testing. All procedures outlined in the plan must be documented and reproducible.

Typical activities include: *model performance evaluation* (applying the model to new data and computing agreed-upon metrics), *error analysis* (examining cases of misclassification or large residuals), and *robustness checks* (for example, testing stability under varied input conditions or simulated noise). A crucial sub-step is verifying that the evaluation data are truly independent of training data to prevent leakage. If the plan included comparisons to benchmark models or baselines, these should be executed. For dynamic models, execution may also involve prospective validation (monitoring performance on live data as the model is deployed).

During execution, any unexpected results or deviations from the plan should be noted. For instance, if performance is far worse than anticipated, sponsors may pause and investigate data issues or model overfitting. In a well-documented process, all raw results, code, and versioning should be archived for transparency. This data analysis is evidence-based: the plan should have established what counts as a success. Did the model meet the pre-specified acceptance criteria? Did uncertainty bounds overlap tolerances? These questions are answered here.

Step 6: Document Results and Deviations

Step 6 requires compiling a **credibility assessment report** that details the outcomes of Step 5 ⁽⁵⁰⁾ www.dlapiper.com. The report should be clear, structured, and comprehensive. It must describe the model's objectives and COU (reiterating Steps 1–2), summarize the validation activities performed, present the performance results (often with tables or plots of metrics), and explicitly note any deviations from the original plan.

Documentation is a fundamental trust-building activity. The report is intended to provide regulators (and internal stakeholders) with evidence that the model meets expectations. As the guidance states, it should describe the results of

all credibility activities and explain any changes in methods or unexpected behaviors (^[50] www.dlapiper.com). For example, if a certain subpopulation was underrepresented in the test data, the report should highlight that shortcoming and its implications. If a hyperparameter was adjusted outside the initial range, that change should be logged.

The guidance leaves “whether, when, and where” to submit this report somewhat flexible – it can be part of a regulatory submission, a meeting package, or simply kept on file ready for inspection (^[50] www.dlapiper.com). Nonetheless, sponsors must plan as if the FDA might request it. An organized, readable report might include sections or appendices for code listings, data schemas, and QA checks. The underlying principle is **transparency**: the more fully one documents, the easier it is for reviewers to assess credibility.

Step 7: Determine Model Adequacy for the COU

The final step is evaluative: **Is the AI model adequately credible for its intended use?** The sponsor reviews the total evidence and decides whether the model’s performance and risk mitigation meet the standards for that COU. If yes, the model can proceed as proposed. If not, the guidance explicitly outlines **five possible follow-up actions** (^[51] www.dlapiper.com) (^[52] www.wcgclinical.com):

1. **Reduce model influence with supplementary evidence.** For example, combine the AI output with additional clinical or manufacturing data so that decisions do not rely solely on the model (effectively lowering model influence).
2. **Increase validation rigor or data.** This means collecting more data or performing deeper tests. For instance, incorporating more training samples or conducting additional external validation to boost confidence in performance.
3. **Add risk-mitigating controls.** Introduce procedural safeguards (e.g. human review checkpoints, alarms) that reduce the consequences of a potential model error.
4. **Modify the modeling approach.** If the existing model family isn’t adequate, try a different algorithm, feature set, or approach altogether.
5. **Re-scope or reject the model.** It may be necessary to narrow the COU, postpone its use, or abandon the model if it cannot be rendered credible.

These options illustrate the iterative nature of credibility assessment. If a model fails to meet acceptance criteria, sponsors do not simply “give up”; they should consider adjustments. For example, if a predictive model for dosing errors proves insufficient on its own, adding an independent laboratory measurement check (supplementary evidence) could down-classify it to a lower influence scenario (^[51] www.dlapiper.com). Conversely, if a high-risk model passes all tests with wide margins, sponsors should still plan life-cycle maintenance.

Life-cycle maintenance. The guidance emphasizes that AI models may change over time (data drift, retraining, etc.) and requires a maintenance plan. This goes slightly beyond the seven steps as written, but it is implied here. If the model is deployed, the sponsor should have in place monitoring schedules, re-validation triggers, and version control procedures (^[47] www.wcgclinical.com) (^[48] www.foley.com). For instance, a machine learning model used in ongoing pharmacovigilance should have processes to detect if its performance degrades on new safety reports. The credibility plan or report should briefly outline these maintenance actions (though some details might remain as internal SOPs).

By the end of Step 7, the sponsor has answered the key question: *Is this AI model sufficiently credible to support the regulatory decision in question?* If not, remedial actions must be taken as outlined. Only when adequate credibility is established can the model’s results be submitted or relied upon in a regulatory context.

Summary of the 7 Steps

The following table summarizes the FDA’s seven-step framework and key activities at each step:

Step	Primary Focus	Key Activities / Evidence
1. Define Question of Interest	Specify the exact decision/addressed by AI.	Articulate the regulatory <i>question or decision problem</i> (e.g. patient risk stratification, QC pass/fail). Document why AI is needed to answer it ([29] www.bioprocessonline.com) ([53] www.foley.com).
2. Define Context of Use (COU)	Describe the AI's role, scope, data, and workflow.	Detail the model's <i>scope</i> , <i>data inputs/outputs</i> , <i>user roles</i> , and <i>how outputs</i> will be used. Specify <i>data-quality criteria</i> and other evidence sources. e.g., "AI predicts vial fill level; findings are reviewed by QC team" ([36] www.bioprocessonline.com) ([24] www.foley.com).
3. Assess Model Risk	Evaluate influence & consequence of errors.	Classify risk by <i>model influence</i> (degree of autonomy) and <i>decision consequence</i> (severity of wrong decision) ([54] www.bioprocessonline.com) ([4] www.wcgclinical.com). Consider probability of error, bias, detection difficulty.
4. Develop Credibility Plan	Plan validation and evidence activities.	Outline model description, data sources, training process, and intended <i>validation strategy</i> appropriate to risk. Specify metrics, test datasets, acceptance criteria, bias checks, etc ([46] www.dlapiper.com) ([47] www.wcgclinical.com).
5. Execute Plan	Conduct testing, validation, analysis.	Perform model training/tuning. Run validation experiments. Calculate performance metrics (accuracy, sensitivity, calibration, etc). Document results of tests, including edge-case and stress tests. Remediate errors.
6. Document Results & Deviations	Record outcomes and any changes.	Compile a credibility report : include performance tables, plots, and narrative on how results compare to planned criteria. Describe any deviations (data issues, changes in model, unexpected findings) and their rationale ([50] www.dlapiper.com).
7. Determine Model Adequacy	Decide if model is fit for COU; or refine.	Evaluate if credibility is met. If yes, proceed. If no, implement mitigations: e.g. add additional evidence, tighten validation, integrate human oversight, revise model or COU as needed ([51] www.dlapiper.com) ([47] www.wcgclinical.com). Include life-cycle maintenance plans.

Each step builds on the last, forming a structured workflow. Together, these steps ensure that by the time an AI model is submitted as part of a drug or biologics application, there is **transparent evidence** that it is trustworthy for its intended purpose. The framework implicitly mirrors traditional risk management: sponsors must define specifications (Steps 1-2), analyze potential failures (Step 3), test and verify (Steps 4-6), and respond to gaps (Step 7) – all in documentation coincident with Good Practices. Throughout, FDA guidance stresses **human oversight and interpretability**: even if an AI model is complex, outputs must be validated and understandable enough to “foster trust in the credibility of AI model outputs” ([55] pmc.ncbi.nlm.nih.gov).

Data, Model Validation, and Evidence Considerations

Beyond the stepwise process above, fully establishing AI credibility involves deep attention to data quality, validation methodology, and performance metrics. We discuss these cross-cutting themes here, drawing on regulatory guidance and scholarly literature.

Data Quality and “Fit-for-Use”

High-quality data are the lifeblood of credible AI models. The FDA framework repeatedly emphasizes using data that are *fit for use* – meaning relevant, reliable, and representative for the COU ([15] pmc.ncbi.nlm.nih.gov) ([4] www.wcgclinical.com). In practice, sponsors should audit their data for completeness, accuracy, and consistency. Outliers and missing values must be handled systematically, and data provenance should be traceable. The NEJM workshop report stresses this: “Fit-for-use data exist across the drug development ecosystem...but often remain siloed... Importantly, the application of fit-for-use data is not absolute...the data quality should be assessed relative to the specific purpose” ([38] pmc.ncbi.nlm.nih.gov).

Sponsors may need to integrate multiple sources: clinical trial results, laboratory assays, real-world data (RWD) from registries or claims, and manufacturing records. Each source has its own quirks. For example, sensor data from a bioreactor can have time lags or variability that must be characterized. The FDA advises specifying *data management procedures* in the plan, including how data will be harmonized, cleaned, and stored. A lack of data (common in rare

diseases) is a challenge; strategies such as federated learning or synthetic data may be relevant, but require careful validation (^[56] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)).

From the literature, we know that sample size and representativeness are critical. Guidelines recommend that “the amount of collected data is sufficiently large for the intended purpose” and ideally pre-specified (^[57] www.nature.com). In fact, formal sample size calculations (e.g. requiring a minimum number of events per outcome) may be needed to ensure statistical confidence. Where existing data are limited, sponsors should consider techniques like learning curves or conservative effect-size estimates (^[57] www.nature.com). If data are biased (e.g. unbalanced patient populations), plans should include mitigation (resampling, equalization) and monitoring for bias impacts (^[58] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) (^[59] www.foley.com).

Overall, data evidence for credibility includes a clear description of data sources, data quality metrics (e.g. error rates, missingness proportions), and justification of data appropriateness. Demonstrating that training and test data align with the deployment context is key. Many guidance reviewers will scrutinize whether the AI encountered “surprise” scenarios. Thus, comprehensive data analysis – from exploratory statistics to documentation of data lineage – is an indispensable part of credibility evidence (Steps 4–6).

Model Validation and Performance Metrics

Once data are established, rigorous validation activities are essential. The guidance does not prescribe specific algorithms, but it expects that whichever modeling techniques are used, they are *thoroughly tested*. Standard metrics (e.g. accuracy, sensitivity, specificity, AUROC) are commonly reported (^[7] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). In internal data analysis, metrics like positive predictive value or F1-score may also be valuable, depending on the regulatory question. For regression or continuous outputs, error measures (MSE, RMSE) and visualizations (e.g. calibration plots) are useful.

Performance metrics should align with the *consequences* identified in the risk assessment. For high-consequence tasks (like patient monitoring), emphasis may be on minimizing false negatives (sensitivity) and quantifying uncertainty bounds. For manufacturing quality control, metrics might include defect detection rates and false alarm rates. The planned acceptance criteria (Step 4) should specify target values or ranges for these metrics, preferably benchmarked against current standards or alternative methods.

Because AI models can overfit, external validation is particularly important. This could involve, for example, holding out an entire clinical trial for testing, or using independent laboratory data for validation. The guidance implicitly encourages all forms of validation (cross-validation, prospective trials, retrospective hold-out data, etc.) to build confidence.

Stability testing is also pertinent. For instance, sensitivity analyses (varying input distributions, adding noise, or simulating worst-case scenarios) can reveal model robustness. One academic review found that while performance monitoring strategies are increasingly discussed, practical guidelines are scarce (^[60] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). This uncertainty underscores the value of the FDA guidance: it signals to sponsors that some form of continuous monitoring (especially for self-learning models) is expected.

In summary, sponsors must collect evidence on model performance that directly ties back to the credibility question. This includes: clear definitions of all calculated metrics, results on test data, interpretations of model errors, and an honest accounting of limitations. In the credibility report (Step 6), such evidence is typically presented in tables/figures. Reviewers will look for, for example, sensitivity/specificity values, confidence intervals, and notes on whether these meet clinical or technical tolerances. Incomplete or cherry-picked reporting will undermine credibility.

Algorithmic Transparency and Interpretability

Although not explicitly labeled as a separate step, **transparency** is a pervasive theme. FDA expects sponsors to provide sufficient description of the model architecture, parameters, and functioning. Even for “black-box” methods like deep neural networks, some countries (and likely soon global norms) insist on explainability logs or surrogate interpretability methods (^[55] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) (^[5] www.foley.com).

The guidance's Appendix (and related documents) suggests disclosing model type (e.g. “random forest with 100 trees and max depth 10”), input features, and important hyperparameters. Sponsors should report how the model was *tuned* (grid search, Bayesian optimization, etc.) and whether pre-trained components or transfer learning were used. If proprietary or third-party models are used, the guidance acknowledges the trade secret issue (^[6] www.foley.com) but still requires that, at minimum, the model's logic and feature uses be summarized in the submission (possibly under confidentiality cover).

Explainability tools (e.g. SHAP values, saliency maps) can provide evidence for transparency. FDA has encouraged (particularly for devices) that model outputs be understandable to users. In the drug context, a sponsor might include a post-hoc analysis showing which input variables most influence the prediction, or include a narrative on how the model arrived at a particular decision. The NEJM report from the 2024 workshop explicitly notes that fostering trust requires allowing end-users to “understand and effectively utilize AI outputs” (^[55] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). Therefore, as part of credibility, studies that correlate model behavior to known science (e.g., a model's attributions match known pathology markers) would bolster confidence.

In practice, the detailed algorithmic documentation may often be too voluminous to submit; instead, sponsors can provide key descriptive elements in the report, and make detailed documentation available for FDA inspection. However, a summary of the model development pipeline – including feature engineering steps, algorithm choice rationales, and data preprocessing – is expected. This requirement aligns with academic guidance that “the growth of complex data-driven models requires careful quality assessment” (^[61] www.nature.com).

Addressing Bias and Generalizability

FDA's framework implicitly addresses fairness and generalizability under risk assessment and validation. Models biased toward a subgroup (e.g. one ethnic group or lab site) are not credible for universal use. Thus sponsors should analyze model performance across relevant subpopulations, particularly if the COU involves a vulnerable group. For example, if an AI predicts adverse event risk in cancer trials, it should be tested for consistency across genders, age ranges, and demographic factors. Any performance gaps should be documented, and possible mitigation (retraining, separate models) considered.

Literature emphasizes this as well: overfitting and sample biases are common pitfalls. The NEJM workshop participants recommended extensive, representative data collection and “testing models across various demographics” (^[58] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). Academic reviews of AI in healthcare note a lack of uniform guidance on bias metrics, but stress that metrics beyond overall accuracy (such as equality of odds or calibration across groups) are needed (^[7] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) (^[8] www.nature.com). Sponsors should therefore include appropriate fairness metrics in their validation plan if applicable.

Generalizability is similarly covered by scenarios like external validation and prospective use. If a model was developed on data from one region or one manufacturing facility, regulators will want assurance it works elsewhere. Demonstrating generalizability could involve multi-center data or retraining on external datasets. A potential angle is *robustness to data drift*: sponsors should describe how the model will be re-assessed if the underlying data distribution changes (e.g. a new patient population or a manufacturing scale-up) (^[47] www.wcgclinical.com) (^[48] www.foley.com).

Tables and Visual Aids

Where possible, tables or figures should be used to summarize evidence. For example, a table of model performance comparing outcomes on training vs test sets (to show absence of overfitting), or a chart of calibration curves, can be very persuasive. The FDA draft guidance itself includes illustrative tables (e.g. performance vs acceptance criteria). Sponsors are encouraged to include succinct visualizations in submissions for clarity.

Below are example tables summarizing risk assessment and innovation needs, illustrating the above concepts:

Risk Factor	Role in Assessment
Model Influence	Determines weight of model output in decision. High influence means model alone makes decisions (higher risk), low means human/instrument checks (lower risk) ([4] www.wcgclinical.com).
Decision Consequence	Impact of wrong decision on safety/quality. High consequence (e.g. patient harm) raises risk rating; low consequence (minor quality flags) reduces it ([4] www.wcgclinical.com).
Data Integrity	Quality of input data. Poor or biased data increases error likelihood (thus risk) [5]. Connections with RWD, lab variation, etc.
Detection Mechanisms	How easily errors are caught. If robust QC exists (good detectability), risk is lower. If error could go unnoticed, risk is higher.
Change Over Time	Likelihood of performance drift. Self-learning or unmonitored models have added lifecycle risk; plans to monitor reduce this risk factor.

(Example: An AI recommending intensive patient monitoring (high influence) that misidentifies low-risk patients could cause severe harm (high consequence) – a high-risk case ([4] www.wcgclinical.com) ([42] www.bioprocessonline.com). By contrast, an AI flagging some vials for human re-inspection is moderate risk because human QC mitigates consequences ([35] www.wcgclinical.com).)

FDA AI Guidance–Driven Innovation Opportunity	Example Implementation
Explainable Models	Incorporate explainable AI techniques that elucidate how predictions are made (e.g. feature importance scores) ([62] www.foley.com).
Bias Detection/Correction	Develop automated tools to identify and correct biases in training data or outputs (e.g. adversarial debiasing) ([63] www.foley.com).
Lifecycle Monitoring Tools	Implement systems to detect data drift and trigger retraining or alerts (e.g. CI/CD pipelines with AI performance tests) ([48] www.foley.com).
Robust Testing Methods	Use synthetic or independent validation datasets to rigorously test generalizability and worst-case performance ([48] www.foley.com).
GMP/CLIA Integration	Build AI workflows compliant with Good Manufacturing/Clinical practices (e.g. version control, audit trails) ([64] www.foley.com).
Automated Documentation	Create software to auto-generate credibility assessment reports capturing versioning, data lineage, and validation summaries ([65] www.foley.com).

Table: Examples of innovations spurred by FDA's transparency and risk requirements ([5] www.foley.com).

Perspectives and Case Studies

The new AI guidance impacts various stakeholders. We examine industry and expert perspectives alongside illustrative case examples.

Industry Interpretation and Adoption

Many in the pharmaceutical industry view the draft guidance as **pivotal**. According to DLA Piper, sponsors now have “a structured framework to assess and address model risk” ([66] www.ark-biotech.com). Firms are recognizing that AI can now **replace certain experiments or manual tasks** in a regulated way. For example, Ark Biotech notes FDA's backing of simulation models “shifting the regulatory path” and enabling companies to “embrace AI with greater confidence” ([67] www.ark-biotech.com) ([68] www.ark-biotech.com). Their industry blog emphasizes early engagement with FDA and tailoring

validation to the stated COU. Similarly, a law firm (Foley) warns clients that **trade secrets will be challenged** by these transparency rules, suggesting sponsors patent AI methods before disclosure (^[6] www.foley.com).

Surveys (e.g. Beroe/ClinicalLeader) indicate significant AI investment in pharma: while exact figures vary, one report suggests *most major companies are now piloting AI in R&D* (with estimates like “69% of pharma orgs invest in AI” commonly cited from industry analyses, though independent verification is limited (^[69] www.allaboutai.com)). Many companies have active AI pilot programs: a recent workshop revealed hundreds of industry participants, with about one-third already using AI in process development or quality control, and many planning submissions involving AI in drug manufacturing within five years (^[70] aapsopen.springeropen.com). In this light, the FDA guidance is seen as timely: stakeholders want clarity on how to present AI in new drug applications.

One recurring industry theme is **risk stratification**: sponsors are categorizing their AI projects as “low-, medium-, or high-risk” under FDA’s scheme, to budget appropriate resources. A DLA Piper alert summarizes this: low-risk projects (e.g. exploratory tools) require basic documentation, whereas high-risk (e.g. pivotal trial models, critical control attributes) will need detailed validation and FDA meetings (^[3] www.ark-biotech.com). Some companies are creating internal “AI quality units” to implement this framework. Early-industry case studies (below) illustrate this transition.

Case Study: AI in Bioprocess Development (Perfusion Rate Example)

Consider a biopharmaceutical process development use case, as described by Ark Biotech (^[9] www.ark-biotech.com): A firm is finalizing a continuous mAb manufacturing process and needs to pick the optimal perfusion rate (a parameter affecting yield and quality). Traditionally several bioreactor runs at different rates would be tested. Instead, the company built a hybrid AI model: it combines a mechanistic process simulator with machine-learning adjustment layers trained on historical bioreactor data. The model **predicts** process outcomes (product titer and quality attributes) across perfusion rates by simulating many scenarios (^[71] www.ark-biotech.com).

Applying the FDA’s 7-step framework:

- **Step 1:** Define the question: “*What perfusion rate yields optimal product quality for Drug X?*” This emphasizes a manufacturing decision tied to product quality.
- **Step 2:** COU: The AI model’s outputs (predicted titer, glycosylation profiles, etc.) will guide selection of the critical parameter for filing. The team specifies that this simulation complements one confirmatory run; historical runs and QC analytics will also inform the final control strategy.
- **Step 3:** Risk: Model influence is moderate (it suggests run at 1.75 VVD, but a validation experiment will confirm). Decision consequence is high (the perfusion rate will be included in the filing and impacts product quality). The model is thus medium-high risk.
- **Step 4:** Plan: The credibility plan includes gathering all historical runs (data) and results of the planned single validation run. It sets performance criteria: the model’s predicted CQA values must match the measured values within predefined acceptance intervals.
- **Steps 5–6:** Execute/Document: The model is run, predicts optimal perfusion 1.75, and an experimental run at 1.75 is performed, yielding product quality data. Results (prediction vs observed) are documented in the report, along with any observed noise or adjustments made.
- **Step 7:** Adequacy: If the model’s agreement with experiment meets the criteria (and no unanticipated biases are found), the model is deemed credible. If not (say predictions were off), the sponsor might adjust the model (additional data) or lower its influence by considering alternate evidence in setting the perfusion rate.

In Ark’s narrative, the framework helped them **integrate AI into regulatory submissions** safely. They note that a regulatory reviewer could ask for additional supporting data or controls if needed – for instance, adding error bars to the

model output to quantify uncertainty (^[9] www.ark-biotech.com) (^[72] www.ark-biotech.com). Importantly, the model was only used to *inform* the decision, not blindly override bench testing; this human-in-the-loop approach lowered perceived risk. Ark suggests that by following the steps, they were ready to include the AI justification in their Biological License Application.

Case Study: AI in Manufacturing (Digital Twin for Deviation Management)

FDA's examples include manufacturing scenarios, and AstraZeneca's process teams have piloted similar ideas. In one illustrative scenario (^[73] www.ark-biotech.com), a continuous bioreactor is equipped with an AI-driven "digital twin" trained on process data. During a production run, a deviation occurs (the nutrient feed is delayed). The digital twin predicts that this delay (within certain bounds) will *not* cause out-of-specification CQAs (glycosylation, aggregation, etc.). Relying on this prediction, quality assurance allows the run to continue with only intensified monitoring, rather than stopping the batch. The actual measured CQAs post-run are indeed within limits, validating the AI's forecast.

Applying the framework:

- The COU is "deviation management" – using AI to decide if a batch still meets specs despite an anomaly (^[73] www.ark-biotech.com).
- Risk is high if the model alone were used, but here human oversight (the batch was examined by QA with the AI recommendation as input) kept it at medium risk.
- The company had planned this scenario with FDA in an early meeting, walking through the 7 steps in advance (^[72] www.ark-biotech.com). They had agreed on what evidence the model needed to generate (e.g. prediction accuracy vs historical deviations). They also agreed any post-call monitoring would be reported per standard deviation protocols.
- After execution, the team documented the successful run in the batch record and noted that this real-world outcome further strengthens (and updates) the model under their Continued Process Verification plan (^[74] www.ark-biotech.com).

This case illustrates two points: (1) Lifecycle use – the model is continuously refined with data from each run, and deviations feed back to its credibility; (2) Regulatory integration – early engagement with FDA on potential AI use in deviation management can smooth later acceptance. Ark's blog and FDA examples both underscore that **contextual controls** (QA oversight, inclusion in batch records) can be used to manage risk while still benefiting from AI insights.

Other Examples and Perspectives

- **Clinical Trial Design.** AI is being tested for adaptive trials, patient screening, and endpoint detection. The CTTI workshop noted uses like optimizing site selection and automating imaging scoring (^[75] pmc.ncbi.nlm.nih.gov). For instance, an AI might predict which trial site will have better enrollment speed (risk to efficacy if wrong). Application of the credibility steps is analogous: the question (boost enrollment), COU (AI ranks sites; humans still approve), risk (wrong site = trial delay), etc.
- **Real-World Evidence (RWE).** AI is used to mine EHR data for safety signals or efficacy trends. Here, Step 3 flags high risk if decisions (e.g. label change) depend on the model. FDA's guidance suggests robust external validation on separate datasets, acknowledging the earlier rule that RWE be "fit for use" in decision-making (21st Century Cures Act lineage).
- **Pharmacovigilance.** AI can scan large post-market databases to detect adverse events faster. This often carries less direct regulatory submission risk (it's about surveillance), but the guidance implies even here, AI outputs should

be credible (i.e. not flagging too many false positives). A “COU” could be defined as generating safety hypotheses for further investigation (lower-risk COU).

From industry and academia, there is widespread support for the guidance’s core tenets. A Future Medicine article notes the draft “emphasizes context of use and a risk-based framework” and requires sponsors to “rigorously test AI models, ensuring accuracy, reliability, and absence of bias” ⁽⁷⁶⁾ www.futuremedicine.com ⁽⁷⁷⁾ www.futuremedicine.com. The NEJM AI workshop report, while predating the guidance by a few months, clearly echoes these principles and explicitly endorses a *flexible risk-based approach* ⁽⁷⁸⁾ pmc.ncbi.nlm.nih.gov. In short, credible consensus is forming: safe AI in drug development needs well-defined questions, robust evidence, transparency, and ongoing oversight ⁽⁵⁵⁾ pmc.ncbi.nlm.nih.gov ⁽⁷⁹⁾ pmc.ncbi.nlm.nih.gov.

Implications, Challenges, and Future Directions

The FDA’s 7-step framework is a major step in the evolving regulatory landscape for AI, but it also raises new questions and opportunities. Here we discuss how the guidance may shape future practices.

Regulatory and Industry Impact

- **Shift Toward Innovation-Friendly Regulation.** This guidance, along with FDA’s AI device policies, signals that regulators are open to AI-enabled approaches – provided they are properly vetted. The structured framework turns uncertainty into a plan: sponsors now have a clearer path to *legitimize* innovative AI methods (like in silico trials or real-time manufacturing control) under FDA review ⁽⁶⁷⁾ www.ark-biotech.com ⁽¹¹⁾ www.foley.com. Over time, this may encourage more AI integration across the drug lifecycle, as companies can anticipate FDA’s requirements.
- **Increased Documentation Burden.** A downside is the higher burden on sponsors to document internal processes. Large amounts of technical detail (data dictionaries, code descriptions, validation protocols) may need to accompany applications. Foley highlights that this “poses a significant challenge for maintaining these innovations as trade secrets” ⁽⁶⁾ www.foley.com. Companies may need to choose between secrecy and compliance: e.g., filing patents on AI methods, or using FDA’s confidential information protections. There is also the practical issue of resource allocation – smaller sponsors may lack the in-house expertise to execute such thorough credibility plans, implying a potential role for AI/consultancy service providers to help.
- **Global Harmonization and Guidance Convergence.** Other jurisdictions are watching, and convergence is already visible. The European Medicines Agency (EMA) issued a *Reflection Paper* on AI in medicines (Sept 2024) emphasizing principles like transparency, data quality, and stakeholder engagement (similar to FDA). Most significantly, on **January 14, 2026** the FDA and EMA jointly released the “*Guiding Principles of Good AI Practice in Drug Development*” — 10 high-level principles covering the entire product lifecycle, including: human-centric design, a risk-based approach, adherence to standards, clear context of use, multidisciplinary expertise, data governance and documentation, model design and development practices, risk-based performance assessment, life-cycle management, and clear essential information ([FDA Guiding Principles](#)) ([McGuireWoods analysis](#)). While the joint principles are non-binding and do not replace the FDA’s draft guidance, they explicitly align with the 7-step credibility framework and foreshadow closer regulatory coordination. Companies developing AI globally will increasingly be able to plan to a shared set of expectations rather than satisfying divergent frameworks.
- **Lifecycle Management and Post-Marketing Surveillance.** The guidance’s model lifecycle emphasis resonates with broader FDA initiatives. For AI, this might align with existing guidelines on change management (for device software) and on pharmacovigilance (for traditional drugs). In practice, we may see sponsors integrate AI model changes into their Quality Management Systems, using mechanisms akin to Post-Approval Changes (PACMP) or Predetermined Change Control Plans (as is done for some software devices) ⁽⁸⁰⁾ www.dlapiper.com ⁽⁸¹⁾ www.dlapiper.com. From a practical standpoint, companies will have to plan ongoing monitoring metrics (e.g. periodic accuracy checks) and triggers for FDA notification when an AI model is updated.

- Academic and Clinical Engagement.** The draft guidance may also spur academic research. Topics like “how much evidence is enough?” or “best metrics for credibility” are still open. Clinical researchers will need education on how to incorporate AI into trial designs and data analysis under this framework. In fact, CTTI and other consortia will likely produce more best-practice documents. We anticipate workshops, white papers, and perhaps FDA Q&A sessions to clarify requirements (e.g. How to choose performance limits? How to validate generative models?). The ultimate test will be how FDA reviewers apply this draft: sponsors should carefully study approval letters and review templates once some AI-inclusive submissions are public.

Tables and Summaries

The FDA guidance and supporting analyses imply useful reference tables. For example, sponsors might use a table mapping **Risk Level** to **Validation Depth**. One can imagine:

Risk Level	Example Use-Cases (COU)	Evidence/Validation Scope
High	Supervising dosing, gating trial access	Extensive: full prospective validation, external datasets, uncertainty quantification, bias audits
Medium	Auxiliary error-flagging, secondary endpoint	Moderate: robust cross-validation, separate hold-out data, partial retraining tests
Low	Informative planning (no patient impact)	Basic: internal validation, limited documentation

Such a table (a form of the examples seen in Foley’s and Ark’s communications) helps allocate review resources.

Another valuable summary tool is a **QA checklist** aligned with the 7 steps. For instance, under each step, a sponsor could list items like “Is question of interest unambiguous and patient-focused?”, “Does COU include data governance details?”, “Have we run an FMEA-style analysis for risk?”, etc. The draft guidance itself can serve as a template for this checklist, but customizing it to a company’s processes would promote compliance.

Challenges and Open Questions

- Regulatory Submissions Practices.** The guidance leaves timing and form of submissions somewhat flexible. Will sponsors be expected to submit the credibility report with every application, or only upon request? The draft suggests discussing “whether, when, and where” with FDA (^[82] www.dlapiper.com), which implies continued uncertainty. This ambiguity can be tricky – companies must decide how much to proactively share. Some may choose to proactively include the plan in an IND package, whereas others might wait for a Biologics License application. The decision may hinge on model risk and novelty.
- Measuring Uncertainty and Explainability.** Deep learning models and large language models (LLMs) are increasingly used in life sciences. The FDA brief mentions Bayesian models favorably for uncertainty quantification (^[44] www.bioprocessonline.com). It remains an open question how regulators will view LLM-based tools (e.g. for line listings or literature review). Will these fall under the same framework, or require a variant? The principles (COU, risk) still apply, but specific metrics for generative models (like hallucination error rates) are still being defined.
- Integration into Quality Systems.** The guidance hints that life-cycle maintenance should be part of a firm’s pharmaceutical quality system (^[80] www.dlapiper.com). In practice, this means CMOs, manufacturers, and sponsors need to establish standard operating procedures for AI. Questions arise: Should AI governance be part of current Computer System Validation (CSV) practices? Who “owns” AI errors? Historically, efficacy and safety liability rests with the sponsor, but when an AI is co-developed by a software vendor, new legal and operational models may emerge.
- Ethical and Social Considerations.** The guidance is primarily technical and science-focused, but it implicitly invokes ethical concerns (bias, transparency, patient welfare). We expect future FDA and HHS discussions on governance frameworks that align with AI ethics principles (e.g. fairness, accountability). Already, FDA’s mention of “responsible and ethical use” in the press release (^[83] www.fda.gov) hints at broader values. Real-world case: if an AI exclusion model wrongly excludes a minority patient group from beneficial therapies, the credibility framework should capture this risk even if outcomes look “accurate” on average. Sponsors must ensure diverse stakeholder input likely in such scenarios.

Future Directions

Looking ahead, several developments may build on this guidance:

- **International Collaboration:** The FDA indicates plans to “engage globally to explore opportunities for harmonizing terminology and basic principles” (^[84] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). We may see joint FDA-EMA workshops or even ICH discussions on AI model validation guidelines. Global pharmaceuticals companies will greatly welcome any convergence.
- **Tool Standardization:** As the framework sinks in, tool vendors and CROs may develop standardized *credibility toolkits* – software to track steps, automate parts of documentation, and even guide metric calculations. Analogous to how bioinformatics has QC pipelines, we may see ML operations (MLOps) adapted to pharma regulation.
- **Case Law and Policy:** If high-profile cases arise (e.g. a mishap blamed on an AI decision), it could spur tighter rules. Conversely, successful real-world uses will build trust. The draft guidance is non-binding (“current thinking”) – stakeholders will watch how it evolves after public comment and see what becomes final.
- **Next-Generation AI:** The guidance addresses traditional ML today. By 2030, new AI paradigms (e.g. quantum ML, autonomous labs, AI-curated data networks) may challenge it. FDA will likely iterate on policy. In fact, FDA has already solicited input on “real-world performance of AI medical devices” (^[85] www.fda.gov), a hint they want to keep updating their approach.

Data Analysis Supporting AI Credibility

To ground the 7-step framework in quantitative terms, we review some key performance considerations and studies:

- **Performance Metrics:** A recent scoping review found that common metrics in clinical AI include Area Under the ROC Curve (AUROC), sensitivity, specificity, and predictive values (^[7] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). These traditional measures are a good starting point, but they may not suffice for regulatory credibility. For example, AUROC summarizes separation across all thresholds, which may hide performance issues at clinically relevant thresholds. Therefore, sponsors should pick metrics that directly tie to clinical impact (e.g. false-positive rate at the operational threshold). In line with standards for predictive models (^[86] www.nature.com), the FDA framework’s credibility plan might also require specifying sample size calculations for validation (e.g. ensuring at least N events for power).
- **Monitoring Over Time:** The ongoing performance of AI models is critical. A 2024 review on monitoring clinical AI noted a dearth of rigorous methods and guidance (^[60] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)), highlighting an evidence gap that the FDA’s emphasis on life-cycle monitoring aims to fill. It found a jump in publications on AI monitoring by 2023, signaling growing interest. Reported monitoring strategies include re-calibration of risk scores, control charts of error rates, and periodic re-training schedules. ADAs (algorithmic drift alarms) might be part of the plan.
- **Uncertainty Quantification:** Bayesian approaches and confidence intervals are gaining attention. For example, if a predicted probability of success is 0.85 ± 0.05 , the sponsor should interpret the ± 0.05 as part of credibility evidence. The guidance encourages characterizing uncertainty “especially for high-risk AI” (^[87] www.futuremedicine.com). This resonates with proposals in literature to express risk model outputs catalytically (e.g. credible intervals for treatment benefit).
- **Bias and Fairness Evidence:** While not explicitly enumerated in the steps, fairness must be demonstrated indirectly by showing consistent performance. If the model is to be used broadly, sponsors should include subgroup analyses (e.g. by age, sex, comorbidity) in the report, and declare any observed performance gaps. This aligns with recommendations from fairness audits in health AI (^[88] www.sciencedirect.com). No publicly available statistics on bias in specific FDA submissions exist, but systematic bias is a known risk (e.g. a 2023 eBioMedicine commentary noted that fairness metrics are still immature in healthcare contexts (^[88] www.sciencedirect.com)).
- **Data Science Best Practices:** Established practices (train/test splits, cross-validation, feature selection, hyperparameter tuning, ensemble methods) underpin rigorous AI. The guidance implies these through terms like “model tuning” and “validation,” but does not forbid novel methods. For example, ensemble or federated learning approaches, often touted in pharma, can be used – but FDA would expect the same credibility steps applied (e.g. if federated learning is used, sponsors would need to explain data sources and how federated training was validated).

In sum, quantitative data analysis is the backbone of steps 5–6. All numerical results, no matter how routine, become evidence. Regulators will expect adherence to recognized model evaluation standards (e.g. TRIPOD reporting, robust cross-validation) even if the guidance itself doesn’t cite them. Therefore, experienced data scientists in pharma should

align their work with established guidelines on predictive model validation, like those reviewed in the literature (^[57] www.nature.com).

Discussion of Implications and Future Directions

The FDA's AI credibility framework brings both opportunities and challenges. We already touched on industry and global perspectives; here we explore longer-term implications and unresolved issues.

- **Regulatory Strategy and IP.** As noted, the requirement for transparency means companies must rethink intellectual property (IP) strategy (^[6] www.foley.com). One implication is a shift from trade secrets (black-box algorithms) toward patents that disclose invention details. Firms may increasingly file patents on AI model structure or training methodology early, to protect themselves while complying with disclosure demands. Legal experts suggest sponsors consider broad patents on the concept and then rely on trade secret only for fully peripheral aspects (^[6] www.foley.com).
- **Quality Management Systems (QMS).** AI credibility will need to be integrated into existing QMS and Quality by Design approaches. For drugs, the ICH Q8-Q10 suite currently covers risk mgmt of processes, but not specifically AI models. We may see new guidance (or updates to existing quality guidelines) that explicitly address algorithmic changes. Companies have begun to include AI governance in their CSV/QMS infrastructure: e.g. defining roles like "AI steward", specifying documentation templates, and including AI topics in audits. Going forward, inspection criteria for AI use will evolve. FDA inspectors may ask to see an "AI POC (proof of concept)" or review records of model performance changes.
- **Harmonization with Medical Device AI.** Many future therapies are drug-device combinations or software-driven devices. The FDA's 2021 guidance on AI/ML in software medical devices (FDA-2021-D-1134) introduced a *Predetermined Change Control Plan* for device AI. While that guidance is device-focused, principles overlap: evaluate safety, set parameters for future modifications, etc. Pharmaceutical AI developers should be aware of this parallel. There could be joint consultations if a drug's AI system has a device component. Harmonizing credibility frameworks across product types will be an interesting policy issue.
- **Impact on Innovation Ecosystem.** Paradoxically, while the guidance imposes controls, experts see it as **pro-innovation**. By delineating a clear pathway, smaller biotech/tech startups might feel more confident pitching AI solutions for regulated drug programs, knowing what evidence to collect. FDA has been actively tracking AI developments: in November 2025 they launched an AI Benchbook and internal courses. This suggests the agency aims to expedite future reviews by upskilling staff now.
- **Patient Safety and Healthcare Delivery.** Ultimately, the credibility framework is about patient safety. All the risk and evidence work is in service of confidence that an AI decision will not harm patients or degrade product quality. If successfully implemented, it could lead to more efficient clinical trials (fewer patients exposed unnecessarily), improved manufacturing robustness (fewer batch failures), and perhaps faster drug approvals due to predictive modeling. In the broader health ecosystem, this framework may serve as a model for how to regulate AI in other contexts (e.g. diagnostic AI or even consumer health AI) with a risk-based, context-specific approach.
- **Evolution of Guidance.** The public comment period on the draft closed on **April 7, 2025**, and the FDA received extensive feedback from industry, academia, patient groups, and software vendors. As of early 2026, the FDA has signaled that **final guidance is expected in Q2 2026**, incorporating comment-period input and aligning with the January 2026 joint FDA-EMA Guiding Principles. Stakeholders had highlighted ambiguities in how to apply the steps to novel AI types — particularly generative AI and large language models, which were barely in scope when the draft was written in late 2024. The FDA will likely clarify issues such as: what constitutes a sufficient 'model risk assessment'; how to handle ensemble models with evolving architectures; how to manage "AI-as-a-service" from external vendors; and whether generative/LLM-based tools require a variant framework. A January 2026 critical review in the *Journal of Chemistry* ([Niazi 2026](#)) flagged the draft's limited treatment of these novel model classes as a key gap the final guidance must address.
- **Ethics and Trust.** Although the guidance is technical, it reinforces trust by insisting on evidence. Patients and practitioners concerned about "black box" AI can take comfort that regulators now explicitly demand demonstration of safety and fairness. In the long run, this may alleviate societal concerns about AI in healthcare, provided sponsors take it seriously. Yet, ethical oversight (beyond what code and data can capture) will still be needed – e.g. explaining AI decisions to patients might require additional tools.

Conclusion

- [8] <https://www.nature.com/articles/s41746-021-00549-7#:~:imple...>
- [9] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance#:~:biore...>
- [10] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:COU%2...>
- [11] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:With%...>
- [12] <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions#:~:The%2...>
- [13] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:or%20...>
- [14] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Artif...>
- [15] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Since...>
- [16] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:leade...>
- [17] <https://www.futuremedicine.com/articles/advancing-the-credibility-of-ai-models-in-drug-and-biologic-development-fdas-proposed-framework#:~:Trans...>
- [18] <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological#:~:This%...>
- [19] <https://regulations.justia.com/regulations/fedreg/2025/01/07/2024-31542.html#:~:estab...>
- [20] <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions#:~:ln%20...>
- [21] <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions#:~:%E2%8...>
- [22] <https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:Al%20...>
- [23] <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions#:~:Al%20k...>
- [24] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:The%2...>
- [25] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:This%...>
- [26] <https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:Al%20...>
- [27] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:FDA%2...>
- [28] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [29] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:The%2...>
- [30] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:ln%20...>
- [31] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:The%2...>
- [32] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [33] <https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:The%2...>
- [34] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:what%...>
- [35] <https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:where...>

- [36] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:Step%...>
- [37] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:monit...>
- [38] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:The%2...>
- [39] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:This%...>
- [40] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [41] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:1,iss...>
- [42] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:For%2...>
- [43] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:shoul...>
- [44] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:essen...>
- [45] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:effec...>
- [46] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [47] [https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:Th...he%2...](https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:Th...)
- [48] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:1,and...>
- [49] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [50] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [51] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:Step%...>
- [52] [https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:1...f%20...](https://www.wcgclinical.com/insights/the-role-of-ai-in-regulatory-decision-making-for-drugs-biologics-the-fdas-latest-guidance/#:~:1...)
- [53] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:Defin...>
- [54] <https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001#:~:Asses...>
- [55] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:indep...>
- [56] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Fragm...>
- [57] <https://www.nature.com/articles/s41746-021-00549-7#:~:It%20...>
- [58] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Selec...>
- [59] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:decis...>
- [60] <https://pmc.ncbi.nlm.nih.gov/articles/PMC11630661/#:~:Concl...>
- [61] <https://www.nature.com/articles/s41746-021-00549-7#:~:While...>
- [62] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:,life...>
- [63] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:decis...>
- [64] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:1,sub...>
- [65] <https://www.foley.com/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance/#:~:,FDA%...>
- [66] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance/#:~:At%20...>
- [67] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance/#:~:The%2...>
- [68] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance/#:~:While...>
- [69] <https://www.allaboutai.com/resources/ai-statistics/drug-development/#:~:%F0%9...>

- [70] <https://aapsopen.springeropen.com/articles/10.1186/s41120-025-00110-w#:~:system...>
 - [71] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance#:~:Inste...>
 - [72] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance#:~:Prior...>
 - [73] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance#:~:Inste...>
 - [74] <https://www.ark-biotech.com/insights/making-pharma-ai-ready-applying-the-fdas-draft-guidance#:~:The%2...>
 - [75] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Key%2...>
 - [76] <https://www.futuremedicine.com/articles/advancing-the-credibility-of-ai-models-in-drug-and-biologic-development-fdas-proposed-framewor#:~:,incl...>
 - [77] <https://www.futuremedicine.com/articles/advancing-the-credibility-of-ai-models-in-drug-and-biologic-development-fdas-proposed-framewor#:~:Respo...>
 - [78] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Works...>
 - [79] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:Speak...>
 - [80] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:,nece...>
 - [81] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:the%2...>
 - [82] <https://www.dlapiper.com/en-ae/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai#:~:,subm...>
 - [83] <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions#:~:The%2...>
 - [84] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12690500/#:~:infra...>
 - [85] <https://www.fda.gov/medical-devices/digital-health-center-excellence/request-public-comment-measuring-and-evaluating-artificial-intelligence-enabled-medical-device#:~:,not%...>
 - [86] <https://www.nature.com/articles/s41746-021-00549-7#:~:clear...>
 - [87] <https://www.futuremedicine.com/articles/advancing-the-credibility-of-ai-models-in-drug-and-biologic-development-fdas-proposed-framewor#:~:Centr...>
 - [88] <https://www.sciencedirect.com/science/article/pii/S2352396423000907#:~:....ma...>
-

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.