IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

FDA Draft Guidance on AI in Drug Development Explained

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-04-01 · fda draft guidance · ai drug development · credibility framework · context of use · pharmaceutical regulation · regulatory submissions · fda ai policy

Original article: https://intuitionlabs.ai/articles/fda-draft-guidance-ai-drug-development



In this article

1. Executive Summary
2. Introduction
3. Historical Background and Context
4. AI in Drug Development: Trends and Challenges
5. Regulatory Framework and Historical Precedents
6. Scope and Applicability of the Guidance
7. In-Scope Uses of AI
8. Out-of-Scope Uses of AI
9. The FDA's 7-Step Credibility Framework
10. Contextual Considerations
11. Defining Credibility in AI Context
12. AI Model Risk is Context-Specific
13. Relation to Other Regulations
14. Data, Model Validation, and Evidence Considerations
15. Case Studies and Example Scenarios
16. Perspectives and Implications
17. Industry Sponsors and Developers
18. Regulatory and Policy Perspective
19. Expert and Industry Commentary
20. Impact on Patients and Public Health
21. Future Directions
22. Conclusion

Executive Summary

In January 2025, the U.S. Food and Drug Administration (FDA) issued its first draft guidance on the use of artificial intelligence (AI) in the development of drugs and biological products (www.fda.gov). Titled *"Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products"* (FDA-2024-D-4689), the guidance establishes a **risk-based, 7-step credibility framework** for sponsors to assess and document the trustworthiness of AI models in a specific *context of use* (COU) (www.fda.gov) (www.dlapiper.com). Critically, the guidance expressly **excludes** certain AI applications: specifically, (1) AI used in **drug discovery**, and (2) AI used to **streamline internal operations** (for example, drafting a regulatory submission) when such use *"does not impact patient safety, drug quality, or the reliability of results"* (regulations.justia.com) (www.dlapiper.com). In effect, the guidance focuses only on AI tools that generate data or insights **directly informing regulatory decisions** about a drug's safety, effectiveness, or quality.

This comprehensive report analyzes the FDA draft guidance in detail. We discuss the historical context of AI in drug development and regulatory submissions, outline the scope and key provisions of the guidance (including in-scope and out-of-scope applications), and explain the draft's proposed 7-step credibility framework. FDA

lists the document as a draft not for implementation, containing nonbinding recommendations (FDA guidance status). Sponsors may use an alternative approach that satisfies applicable statutes and regulations (FDA draft). We draw on FDA materials, industry analyses, and expert commentary to present multiple perspectives. The report includes case scenarios (both hypothetical and real-world) to illustrate how AI is being used today in pharmaceutical R&D and manufacturing, and how sponsors should prepare to meet the new FDA expectations. We also examine the implications of the guidance for industry innovation, regulatory strategy, and patient safety, and consider future directions as AI technologies advance. All claims are substantiated with citations to the guidance, FDA announcements, regulatory analyses, and relevant literature.

Introduction

Artificial intelligence and machine learning (AI/ML) technologies are transforming many aspects of healthcare and drug development. Over the past decade, sponsors have increasingly incorporated AI tools to analyze complex datasets, optimize clinical trials, and predict patient outcomes. These tools can accelerate drug discovery, improve manufacturing processes, and generate real-world evidence from health records. However, AI's rapid rise also raises new regulatory challenges. Chief among these is **credibility**: regulators must be able to trust that an AI model's output is accurate, reliable, and appropriate for its intended regulatory role. Without clear guidance, companies have faced uncertainty about what AI information the FDA will accept, and what documentation will be required in submissions.

The FDA has recognized the urgent need for clarity. In January 2025, FDA Commissioner Robert Califf announced a draft guidance to address "the use of artificial intelligence (AI) intended to support a regulatory decision about a drug or biological product's safety, effectiveness or quality" (www.fda.gov). This was the agency's **first guidance on AI in drug development**. The guidance provides *non-binding recommendations* on how to establish and demonstrate an AI model's credibility for regulatory decision-making. It builds on FDA's previous experiences: over 500 drug or biologics submissions with AI components have been received by the FDA since 2016, particularly in areas like oncology and neurology (www.statnews.com). As expected, FDA has seen an exponential increase in AI uses in submissions since 2016 (www.fda.gov), and has collaborated across its centers (CDER, CBER, CDRH, CVM, OCP, and OII) to craft a unified framework.

Key points of the FDA draft guidance include:

- **Scope (Included Uses):** AI applications across the drug lifecycle are covered if their outputs produce data or information *intended to support regulatory decisions* about a product's safety, efficacy, or quality (regulations.justia.com). This encompasses uses in nonclinical research, clinical trials, manufacturing, and post-market surveillance. For example, using AI to select patients for trials, optimize dosage based on model-informed pharmacology, detect safety signals in real-world data, or control manufacturing quality would fall under the guidance (www.foley.com) (www.goodwinlaw.com).
- **Scope (Excluded Uses):** The guidance **does not** apply to certain AI uses that *do not directly affect a regulated output*. Specifically, it *excludes* AI used **for drug discovery** and AI used to **streamline internal operations** (such as drafting regulatory documents, managing workflows, or allocating resources) when those tasks do not impact safety, drug quality, or study results (regulations.justia.com) (www.dlapiper.com). The exclusion concerns this draft's scope: operational uses such as drafting a

submission are excluded when they do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). It does not establish a general exemption from applicable statutes or regulations.

- **Risk-Based Credibility Framework:** The guidance introduces a **7-step framework** for establishing the credibility of an AI model in its specific *context of use* (www.goodwinlaw.com) (www.dlapiper.com). "Context of use" (COU) refers to the precise role of the model – what question it addresses and how its output will be used. The steps range from defining the question and COU, to assessing model risk, developing and executing a credibility assessment plan, and finally determining if the model is adequate for its use (www.goodwinlaw.com). The draft recommends a framework for establishing and evaluating model credibility for a particular context of use; these are nonbinding recommendations (FDA guidance status).
- **Regulatory Interaction:** The draft guidance encourages early and ongoing engagement with FDA. Sponsors are urged to meet with the agency **before** using a high-risk AI model in a critical decision, to align on expectations for the credibility plan (www.goodwinlaw.com) (www.dlapiper.com). The FDA does not necessarily require submission of all credibility documentation, but sponsors must decide whether and when to share the plan and evidence with the agency. The agency will consider how the AI was validated when reviewing applications (INDs, NDAs/BLAs, etc.) containing AI-derived data, and may request further justification if needed.

This guidance is situated in a broader policy context. The European Medicines Agency (EMA) issued its own *Reflection Paper* on AI in drug development in September 2024 (www.statnews.com). While FDA focuses explicitly on credibility and risk in decision-making, EMA's approach also addresses principles and good practices for AI use across the lifecycle (www.statnews.com) (www.ema.europa.eu). Industry experts note that "regulatory clarity is one of the top three barriers" to wider AI adoption in pharma (www.statnews.com), so the FDA guidance is seen as a step toward removing uncertainty. Meanwhile, AI-enabled medical devices have been governed under separate FDA policies; indeed FDA issued a device AI draft guidance on the same day (Jan 7, 2025) (www.americanbar.org). Taken together, these developments signal that the FDA is integrating AI into its regulatory oversight, treating it as an increasingly important part of the "regulated stack" for drug products.

This report provides an in-depth analysis of the FDA's AI-in-drug-development draft guidance. We begin by detailing the historical and regulatory background of AI in life sciences. We then dissect the guidance's contents: scope, 7-step framework, risk and credibility requirements, and affected stakeholders. We intersperse discussion with case examples (both hypothetical and actual) to illustrate how the guidance applies. We compare the FDA's approach with international perspectives and discuss emerging trends. Throughout, we cite extensive evidence from FDA documents, legal and industry commentaries, and relevant scientific literature to support our examination. The goal is to give readers a complete, authoritative understanding of this guidance and its implications for the future of AI in drug development.

Historical Background and Context

AI in Drug Development: Trends and Challenges

The pharmaceutical industry has long grappled with high costs and long timelines for bringing new drugs to market. Recently, AI and machine learning have offered tools to mitigate these challenges. Machine learning algorithms can sift through complex datasets – genomic data, clinical trial results, electronic health records, literature – to find patterns humans might miss. For example, AI models have been used to predict patient responses to therapy, identify potential new drug targets, optimize chemical synthesis routes, and detect adverse event signals from post-market data (www.fda.gov) (talglaw.com). One study noted that AI is being applied across the drug lifecycle to reduce reliance on animal studies, build predictive clinical trial models, and integrate heterogeneous data sources (talglaw.com). In oncology and rare diseases, sponsors are experimenting with AI-driven patient stratification to enrich clinical trials or adapt dosing dynamically.

Despite these innovations, integrating AI into regulated drug development has been cautious. Regulators and industry have raised concerns about data quality, model bias, interpretability, and robustness over time. AI models can inherit biases in their training data (e.g. underrepresentation of certain populations) or degrade when deployed under new conditions (known as *model drift*). The black-box nature of some AI, especially complex neural networks, also challenges validation: it may be hard to trace *why* an AI made a certain decision. As Ismail Amin notes, key challenges include “variability in data quality, potential biases, difficulty in understanding AI models, and changes in model performance over time” (talglaw.com). In short, regulators worry about trusting AI outputs without adequate evidence.

Historically, companies have often incorporated AI-based analyses under existing regulatory frameworks without specific AI rules. For instance, model-informed drug development (MIDD) tools – such as pharmacometric models to predict a safe starting dose – are not new. The FDA has accepted computational models for certain bridging data. But AI’s ability to “learn” from data dynamically is a new variable. Until recently, most FDA guidance addressing AI focused on medical devices (e.g. draft guidance on AI-enabled SaMD in 2021) or on general medical software. There was no dedicated FDA guidance on AI in drug development. This left ambiguity: if a sponsor used an advanced AI algorithm to analyze trial data or optimize manufacturing, it was unclear what explanation or validation the FDA would expect. Indeed, industry leaders have cited **lack of regulatory clarity** as a top barrier to AI adoption in drug R&D (www.statnews.com).

The pace of AI innovation accelerated with the rise of big data in healthcare and the availability of powerful computational resources. Concurrently, regulators around the world began taking notice. In 2024, the European Medicines Agency (EMA) published a Reflection Paper on AI in the medicinal product lifecycle (www.statnews.com) (www.ema.europa.eu). EMA’s document (finalized in September 2024) discusses principles for AI use *from discovery to post-authorization*, advising developers on good practices and noting potential benefits and risks. The EMA reflection paper emphasizes the need for transparency, validation, and human oversight in AI-based systems. Similarly, in the UK the Medicines and Healthcare Products Regulatory Agency (MHRA) has released guidance on AI use in medical devices and health products. However, until the FDA guidance, no U.S. agency had provided specific recommendations for AI in drug development.

In the U.S., the FDA had been receiving increasing numbers of submissions containing AI. As of late 2024, the FDA reports over *500 submissions with AI components* since 2016 (www.statnews.com). These submissions span numerous therapeutic areas, reflecting a broad interest in leveraging AI (especially in oncology, neurology, and gastroenterology) to improve trial efficiency or data analysis. FDA officials observed an "exponential" rise in AI usage (www.fda.gov). This trend caught the agency's attention: Commissioner Califf stated that with "appropriate safeguards in place, artificial intelligence has transformative potential to advance clinical research and accelerate medical product development to improve patient care" (www.fda.gov). But that potential must be balanced with standard regulatory priorities. The FDA's mission to protect public health demands confidence that decisions driven by AI are valid and reproducible.

Thus, in **January 2025** the FDA proactively issued draft guidance to begin filling the void. This draft guidance – its first for drug-related AI – formally acknowledges the growing role of AI and defines how FDA intends to evaluate it. It provides a structured, science-based approach so that sponsors can prepare appropriate documentation. As Fahimeh Mirakhori (Ph.D.) summarizes: the framework is "structured, risk-based," focusing on COU and associated risks, and guiding both regulators and developers on when AI models can be "trusted to support critical decisions" (www.bioprocessonline.com). By laying out the expectations now, the FDA aims to encourage innovation while ensuring rigorous standards are met.

Regulatory Framework and Historical Precedents

Before this guidance, the FDA's oversight of AI in drug development fell under existing regulations for data and analysis. For example, the medical-device definition is in FD&C Act Section 201(h)(1), 21 U.S.C. §321(h)(1) (FDA device definition; FDA draft, footnote 6). Some AI tools (like an algorithm diagnosing disease from medical images) are regulated as devices, whereas tools purely for research might not be. In the drug context, if an algorithm outputs data that ends up in a New Drug Application (NDA) or Biologics License Application (BLA), the agency would review that as part of the evidence of safety/effectiveness, but historically with no tailored guidance on how to validate the algorithm.

FDA's recognition record for **ASME V&V 40-2018** (recognition number 5-122, entered January 14, 2019) concerns physics models and expressly excludes statistical models used in AI (FDA recognition record). The AI draft draws on the standard's high-level concepts for defining the question of interest, context of use, and model risk; that conceptual use does not extend the recognition to statistical AI models (FDA draft, footnote 13).

FDA's *M15 General Principles for Model-Informed Drug Development* finalizes a draft of the same title issued on December 30, 2024 (FDA M15).

The statutory clinical decision support (CDS) exclusion and this draft's scope are separate questions. FDA explains that a CDS software function must meet all four criteria in section 520(o)(1)(E) of the FD&C Act to fall outside the device definition: restrictions on processing medical images and specified signals; use of medical information; recommendations to a health care professional; and a design that enables that professional to independently review the basis for recommendations rather than rely primarily on them (FDA's CDS criteria). Separately, the AI draft does not address operational-efficiency uses that do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). The draft does not codify statutory exemptions.

Finally, on the machine learning side, FDA's discussion paper, *"Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device,"* dates to April 2019 (FDA action-plan announcement). The draft discusses life cycle maintenance separately in Section IV.B. Step 6 concerns documenting credibility assessment results and deviations from the plan (FDA draft, Sections IV.A.6 and IV.B).

In summary, the FDA's 2025 draft guidance builds on these precedents by explicitly addressing **AI/ML models used for regulatory support in drugs**. It sets expectations that sponsors define clear contexts of use for their models, assess risks, and compile credibility evidence. The guidance is careful to distinguish AI uses that *should* be part of regulatory review versus those that are outside. The remainder of this report delves into the language and implications of the guidance, step by step.

Scope and Applicability of the Guidance

The draft guidance clearly delineates which AI uses are *in-scope* and which are *out-of-scope*. Understanding this delineation is crucial for sponsors to know whether their AI activities will attract FDA scrutiny under this guidance.

In-Scope Uses of AI

The guidance applies to any AI model used **during the lifecycle of a drug or biological product** where the AI's output is intended to support a regulatory decision about the product's **safety, effectiveness, or quality** (regulations.justia.com) (www.fda.gov). This is a broad inclusion of virtually all drug R&D stages where AI-generated data can influence outcomes. Specifically, the guidance mentions use during:

- **Nonclinical (preclinical) development**: For example, an AI model that predicts toxic risk from in vitro data or an animal study design. If the AI's output informs the safety profile used in an Investigational New Drug (IND) application, it's in scope.
- **Clinical development**: For example, an AI model that recommends patient inclusion/exclusion criteria, or one that predicts efficacy endpoints from trial data. An AI algorithm that classifies patient risk and thus determines dosing or monitoring during a trial is in scope. FDA's Goodwin analysis gives a hypothetical: an AI model categorizing patients by risk of life-threatening side effects, used to decide outpatient vs. inpatient monitoring. Because an AI error could lead to a fatal outcome, this high-stakes use clearly requires credibility assessment (www.goodwinlaw.com).
- **Manufacturing and Chemistry, Manufacturing, and Controls (CMC)**: AI tools that optimize manufacturing parameters, detect out-of-specification batches, or ensure product consistency would be in scope if their outputs affect product quality. For instance, an AI system that checks quality control (QC) assay data and flags contamination risks for action is within the guidance. A low-risk example from DLA Piper: an AI identifies manufacturing batches out-of-spec, then a human reviews and executes a corrective action plan (www.dlapiper.com). Even though human oversight exists, because AI influenced QC, the guidance would apply (with an appropriately low-risk credibility plan).

- **Postmarketing surveillance (pharmacovigilance and real-world evidence)**: If AI analyzes electronic health records or adverse event data to generate insights used in regulatory filings (e.g. a periodic safety update report), this is in scope. The FDA explicitly mentions pharmacovigilance contexts and real-world data uses (www.dlapiper.com).
- **Product lifecycle analysis**: The guidance's title and summary confirm it covers **all phases** (preclinical through postmarket) as long as the AI influence is on regulated attributes (regulations.justia.com).

In all these cases, the AI's role is to *produce information or data used for decision-making* about a regulated product. In short, if a sponsor is relying on AI-derived results as part of evidence (e.g. a model-predicted endpoint, simulation, or biomarker analysis) that will appear in an IND, NDA, BLA, or other regulatory submission, the AI is likely in scope.

Out-of-Scope Uses of AI

Crucially, the guidance is **explicit that many AI uses are not covered**. It *excludes* AI used in purely research or operational contexts **unless the AI output directly affects a regulatory decision**. Specifically, the guidance **"does not address"** AI when used in:

1. **Drug discovery**: This includes AI for target identification, lead optimization, virtual screening, and other pre-development activities. For example, if an AI algorithm helps identify a new molecular scaffold before any preclinical tests, that use is out-of-scope. The guidance's language is unequivocal: it "does not address the use of AI models … (1) in drug discovery" (regulations.justia.com). (However, as discussed later, if an AI in "discovery" actually generates data influencing safety/efficacy decisions, say by predicting toxic metabolites, that narrow aspect might tangentially be in scope. But general discovery activities are excluded.)
2. **Operational efficiencies/internal workflows**: The draft does not address operational-efficiency uses, including internal workflows, resource allocation, and drafting or writing a regulatory submission, when they do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). The scope conditions apply to the particular use; medical writing and transcript analysis are not categorically excluded.
3. **"Low-impact" safety/quality tasks**: If an AI influences processes but the end result does not impact patient safety, drug quality, or study reliability, it's out. For example, an AI that automates labeling artwork for a non-critical change might be exempt, unless labeling content itself becomes a regulatory issue. The Federal Register notice explicitly defines the exclusion as any internal use *"that do not impact patient safety, drug quality, or the reliability of results from a nonclinical or clinical study"* (regulations.justia.com).

These exclusions mean the guidance is deliberately **narrow**: it focuses on AI that has *material regulatory effect*. This is consistent with policy goals. FDA is not trying to police whether your company uses ChatGPT to write emails or to optimize office staffing; it's only concerned when the AI's output could influence a decision about whether a drug is safe and efficacious or of acceptable quality. As one summary put it, the guidance is

limited to models that impact core regulatory considerations (natlawreview.com). In practice, sponsors can continue to leverage AI for internal work (often an enormous efficiency gain) without needing to go through FDA's credibility process, as long as those tools are segregated from the regulated evidence.

Table: In-Scope vs Out-of-Scope AI Uses

Application Area	In Scope?	Example
Drug discovery (target ID, lead optimization)	**No:** AI used only to identify potential drug targets or optimize lead compounds before preclinical development.	AI algorithm suggests a novel compound binding a cancer target (no direct safety/efficacy data yet) – *out-of-scope*.
Nonclinical modeling (toxicity, PK/PD)	**Yes:** If AI models generate safety/efficacy data that inform Investigational New Drug (IND) applications. Requires credibility.	AI predicts organ toxicity from molecular structure, informing IND risk assessment – *in scope*.
Clinical trial design & analysis	**Yes:** AI used to set inclusion criteria, predict endpoints, or analyze trial data affecting study outcomes.	AI classifies patients by biomarker, selecting trial participants; or AI method predicts efficacy endpoint based on interim data – *in scope*.
Manufacturing and quality control	**Yes:** AI tools used for process control, batch release decisions, or QC data analysis that affect product quality.	AI-driven monitoring flags a batch outside specifications and suggests corrective action before market release – *in scope*.
Pharmacovigilance and real-world evidence	**Yes:** AI systems mining safety databases or EHRs to generate signals or evidence used in regulatory submissions.	AI identifies an unexpected adverse event pattern from real-world data; findings included in a risk-management plan – *in scope*.
Regulatory writing and review	**Excluded only when:** AI used for operational efficiency in drafting or writing a regulatory submission does not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results.	Human review or the absence of safety-data generation alone does not establish that the use meets all three scope conditions (FDA draft, Section II).
Internal administrative tools	**No:** AI for scheduling, resource allocation, document formatting, etc., where outputs do not influence clinical or quality decisions. → *excluded*.	AI schedules clinic appointments or optimizes lab workflow, but does not affect any patient safety or product quality decisions – *out-of-scope*.
Use of AI without human oversight in high-risk areas	**Yes:** If AI makes autonomous decisions on safety/efficacy without human review, it's higher risk and definitely in scope.	AI independently classifies an MRI scan to include a patient in a trial (no human review) – *in scope and high risk*.
Low-risk oversight (human-in-loop)	**Yes (but lower risk):** If AI provides a recommendation that is reviewed by a human for final decision, it is still in scope (though credibility requirements may be lighter).	AI suggests batch off-spec; human pharmacist reviews and confirms before rejecting the batch – *in scope, with moderate risk* (www.dlapiper.com).

Table 1: Examples of AI applications in drug development and whether they fall under the FDA draft guidance. In-scope uses involve AI outputs that directly feed into regulatory decisions about safety, efficacy, or quality. Out-of-scope uses are for discovery or internal tasks without direct regulatory impact (regulations.justia.com) (www.dlapiper.com).

The table illustrates the line drawn by FDA: AI in discovery and back-office is out, AI used as part of validated development or review processes is in. It is important to note that in-scope vs out-of-scope is judged on the *impact* of the AI's output, not the technology itself. For example, the same natural language AI could draft a private memo (out-of-scope) or assist in writing a patient consent form that is used in an IND (in-scope if content affects risk assessment). Similarly, an AI used in initial lead discovery (out-of-scope) might later be repurposed in a model predicting clinical response (in-scope). The guidance instructs sponsors to *define their context of use precisely* and assess whether the model will affect regulatory conclusions (www.fda.gov) (www.dlapiper.com). Sponsors uncertain about whether an AI application is within scope are advised to engage FDA early for clarification (regulations.justia.com) (www.dlapiper.com).

The FDA's 7-Step Credibility Framework

At the heart of the draft guidance is a **7-step, risk-based credibility framework** for AI models. The framework is meant to be a planning and evaluation process: it guides sponsors on how to build and maintain evidence that their AI model can be trusted ("credible") for its intended use. Each step corresponds to a key element of demonstration. These steps reflect model risk management principles (similar to those in software validation or quantitative model validation) tailored to AI/ML.

FDA identifies these steps as **Defining the context and model**, **Risk assessment**, and then implementing and verifying a **Credibility Assessment Plan**. The steps are:

1. **Define the Question of Interest (QI)** – the problem or decision addressed by the AI model.
2. **Define the Context of Use (COU)** – the precise role and scope of the AI model in answering the QI.
3. **Assess Model Risk** – evaluate the risk that the AI model's use will lead to incorrect decision or outcome.
4. **Plan Credibility Assessment Activities** – develop a systematic plan of tests/analyses to establish model credibility for its COU.
5. **Execute the Plan** – carry out the validation and evaluation activities.
6. **Document Results** – compile a credibility assessment report, including results, conclusions, and any deviations from the plan.
7. **Determine Adequacy** – conclude whether the evidence supports using the model for the COU, or if further work is needed.

Each step is described in detail in the draft and accompanying analysis (see *Table 2*). These steps must be proportional to risk: high-risk applications (where AI model influence on a critical decision is large) demand more thorough evidence. The FDA emphasizes early planning: sponsors should incorporate these steps into their development of AI tools, and engage with FDA to align on expectations at steps 4–5. Although not every step must be submitted to FDA automatically, this framework documents a complete lifecycle understanding.

The 7-step framework, as summarized by Goodwin Law, is:

Step	Action	Key Focus
1	**Define the Question of Interest.**	Identify the specific decision or parameter the AI will address. (www.goodwinlaw.com). Example: “Classify patient as high/low risk for toxicity.”
2	**Define the Context of Use (COU).**	Specify how the model will be used and combined with other information (www.goodwinlaw.com) (www.dlapiper.com). E.g., “Model output will determine hospital monitoring level.”
3	**Assess Model Risk.**	Evaluate risk by (a) *model influence* (extent model affects decision) and (b) *decision consequence* (impact if wrong) (www.dlapiper.com). Example: high influence + high consequence = high risk.
4	**Develop Credibility Plan** – Document the planned activities to demonstrate reliability.	Outline evidence needed: model development details, data sources, validation datasets, performance criteria (www.dlapiper.com). E.g., “Plan to test model on independent trials data.”
5	**Execute Plan.**	Perform the validation activities (e.g. testing, cross-validation, sensitivity analyses). Engage FDA to confirm approach as needed (www.dlapiper.com).
6	**Document Credibility Results.**	Prepare a report with methodology, test results, consistency checks, and any deviations. The report is meant to *"provide information that establishes the credibility of the AI model for the COU"* (www.dlapiper.com).
7	**Conclude Adequacy.**	Decide whether the model is fit for the COU; if evidence is insufficient, then take steps (e.g. collect more data, reduce model influence) to meet requirements (www.dlapiper.com).

Table 2: FDA’s 7-Step Credibility Assessment Framework for AI Models (from draft guidance). Steps involve defining the question and context, assessing risk, planning and executing validation, and concluding adequacy. (Source: FDA Goodwin summary (www.goodwinlaw.com) and DLA Piper analysis (www.dlapiper.com).)

Below we walk through these steps, elaborating on their intent and components. References in brackets cite FDA’s guidance or expert analysis of the guidance.

Step 1: Define the Question of Interest

Each AI application must start with a **clear problem statement**. The guidance calls this the “question of interest” (www.goodwinlaw.com). This is not the same as the COU; it is the specific goal the model aims to achieve. For example, a question of interest might be: *“Can this AI model predict which patients will experience life-threatening side effects from our drug candidate?”* or *“Can the model determine if a manufacturing batch meets quality specifications?”*. Defining the question ensures the model development is purpose-driven, and it helps identify what data and metrics are relevant. The question might come from any point in the lifecycle: in a clinical trial, it could be “should we dose-adjust a patient based on their predicted risk?”; in postmarket, it could be “does the model detect previously unreported adverse events in EHR data?”.

The **sponsors** should document this question explicitly. It sets the stage for all later steps. A poorly defined question leads to a mismatch between model output and regulatory need. The guidance uses examples like patient selection criteria or risk stratification in trials (www.foley.com), but any AI-driven decision point ought to begin with Step 1. As DLA Piper notes, Step 1 requires *“the proposal should define the specific question, decision, or concern being addressed by the AI model.”* (www.dlapiper.com).

Step 2: Define the Context of Use (COU)

After formulating the question, the next step is to specify the **context of use (COU)** of the AI model (www.goodwinlaw.com) (www.dlapiper.com). Context of use is a term the FDA has borrowed from medical device regulation (and from modeling standards). It means: **What role will the model play, and under what conditions?** The COU includes details such as:

- **Scope of the model:** What exactly is being modeled (e.g. patient survival probability given biomarker data)?
- **Use in workflow:** How will the output be used (as a primary decision-maker, a decision support, an alert, etc.)? Will human experts verify it, or act autonomously? (For example, is it fully automated or part of a human-in-the-loop process (www.dlapiper.com).)
- **Data inputs and boundaries:** What inputs does the model use (clinical trial data, lab results, imaging, etc.) and from what population? Are there other information sources that will supplement the model's output (e.g. will clinicians still require an MRI or lab test in addition to the AI output)?
- **Operating parameters:** Will the model ask new queries over time, or is it a one-time analysis? Is it deployed in a closed setting (e.g. within a single trial), or broadly across different studies?

The COU essentially describes the environment and assumptions for the AI model. It is critical because an AI model's credibility cannot be assessed without knowing *how* it is supposed to be applied. The draft guidance stresses that different COUs can lead to different validation needs: an AI in one trial may be high-risk, while the same model in a different, less-critical setting might be lower risk (www.goodwinlaw.com) (www.dlapiper.com). Defining COU also aligns with how FDA reviewers think: they evaluate a model's output only within the stated scope. If sponsors later want to expand the COU (e.g. apply the model to a different patient population), they would likely need to repeat some steps.

In practice, sponsors should answer questions like: *"What exactly is the AI predicting? How will this prediction be used by the trial or manufacturing process? In what patient population or manufacturing context? Is the model being used standalone or together with other data?"* For instance, if an AI model predicts patient response, will doctors automatically adjust doses based on it, or will they see the AI's suggestion and then decide? These details belong in Step 2. A well-defined COU sets the risk level and guides later steps.

Step 3: Assess the AI Model Risk

With the scenario framed, Step 3 is to **assess the risk posed by the AI model** in that context. The FDA guidance recommends evaluating two factors: **model influence** and **decision consequence** (www.dlapiper.com).

- **Model Influence:** How much does the AI model *influence* the decision? If the AI's output is used as a final determinant, influence is high. If it's one of several inputs, a secondary check, or is overridden by human check, influence is lower. For example, an algorithm that automatically determines patient dosage without review is high influence. An algorithm that suggests dosing but requires a doctor's sign-off is moderate influence.

- **Decision Consequence:** What is the potential harm (or benefit) if the model is wrong? This considers the severity of a wrong answer. A decision consequence is high if a mistake could lead to serious patient harm or product failure. For example, misclassifying a high-risk patient as low-risk (preventing necessary care) has high consequence. Misclassifying an expected mild side effect has lower consequence.

Combining these two dimensions yields overall **model risk**. DLA Piper explains: “AI models and use-cases in which AI is used to make a final determination … without human intervention will be considered as higher risk,” especially when safety is involved (www.dlapiper.com). Conversely, models that only support minor decisions with human oversight may be lower risk.

Model risk reflects the combination of model influence and decision consequence in the stated context of use. In FDA’s manufacturing example, high decision consequence combined with low model influence after release-testing mitigations yields medium model risk (FDA draft, Step 3). The draft guidance suggests this assessment should determine how rigorous the subsequent credibility activities must be (www.dlapiper.com).

For illustration, consider two scenarios:

- **High Risk:** An AI model that autonomously classifies whether a patient with a new therapy needs life-saving intensive care. Here model influence is effectively 100% (the AI makes the call), and consequence is high (a wrong call could be fatal). This demands thorough validation of the AI’s performance, as any error has grave implications.
- **Lower Risk:** An AI tool that identifies potential outliers in manufacturing data, but only flags them for human review. Influence is moderate (a human ultimately decides action) and consequence is moderate. Such a scenario still requires validation, but the bar may be somewhat lower. FDA notes that requiring human review *after* the model output generally reduces risk (www.dlapiper.com).

In Step 3, sponsors should document their reasoning on model risk. This includes describing the worst-case impact and how much they are relying on the AI. Some experts suggest using frameworks like NASA’s model credibility assessment or cross-industry risk assessments to quantify this. The key is that risk informs the **stringency of Steps 4–6**: higher-risk models need more data, more testing, and possibly external studies, whereas low-risk models might need only basic internal validation.

Step 4: Develop a Credibility Assessment Plan

Step 4 is to **plan the credibility activities** that will establish the AI model’s trustworthiness for its context of use. Think of this as creating an audit plan for the model. The plan should list the specific tests, analyses, and documentation that will be done (or have been done) to verify the model.

Key elements of the plan include (www.dlapiper.com):

- **Model Description:** Detail of the AI model’s structure (algorithm type, architecture, hyperparameters). For a trained ML model, include how it was built.
- **Data Sources:** Comprehensive description of the data used for training, tuning, and testing. This includes data provenance, preprocessing steps, and any limitations (e.g. missing data, imbalances). For example, if training data came from a particular demographic or region, note that.

- **Performance Metrics:** Identify the metrics and acceptance criteria to be evaluated. This could be accuracy, sensitivity, specificity, area under the curve (AUC), or other domain-specific measures (e.g. positive predictive value for a safety marker). The plan should state performance targets that are acceptable for the COU.
- **Validation Approach:** Specify how the model will be tested. This might involve cross-validation, use of an independent test set, simulation studies, or prospective validation. It should also include robustness checks (e.g. testing on edge cases, stress testing with noisy or unexpected inputs).
- **Comparators:** If applicable, describe any baseline or comparator models. FDA will want to know how the AI compares to existing methods or to standard-of-care.
- **Provisions for Updates:** If the model will be updated (adaptive learning), include how future data will be incorporated and how changes will be tested. (See Step 7 discussion).
- **Data Integrity and Security:** Outline how data was handled ethically and securely, and how the model's code is controlled. (Although not emphasized in the draft, good practice is to maintain version control and audit trails for the model.)
- **Bias and Fairness Checks:** Since bias is a known issue, the plan should consider how fairness will be assessed. For example, analyzing performance across subgroups. The guidance itself doesn't enumerate this, but many commentaries advise it.

The plan is essentially a protocol for model validation. It should be tailored to the COU and commensurate with the model risk. For a high-risk model, Step 4's plan would be extensive: numerous evaluation datasets, rigorous performance thresholds, and possibly third-party review. For a low-risk model, the plan might be simpler.

From a regulatory standpoint, the plan provides transparency: it shows FDA how the sponsor intends to assure quality of the AI. Importantly, the draft guidance indicates that sponsors should decide whether and when to submit the plan to FDA. If the model is critical and/or novel, sponsors might include the plan with an IND or BLA as part of their submission package or discuss it in a pre-sub meeting. If the model is lower risk, they may keep it internally but still share summaries if requested. The FDA states it will "set expectations regarding appropriate credibility assessment activities" during early engagement (www.dlapiper.com).

Step 5: Execute the Credibility Plan

Having a plan, Step 5 is **execution**. This means carrying out all the tests and evaluations as described. The deliverables of this step are the raw results: the model outputs, error rates, statistical analyses, etc. Essentially, this is where the sponsor gathers the evidence. Key activities might include:

- **Testing on Independent Data:** Use held-out data or prospective data (if available) to test how the model performs outside its training set.
- **Sensitivity and Specificity Analysis:** Evaluate how the model's performance metrics behave under different scenarios.
- **Stress Testing:** Challenge the model with edge cases, noisy inputs, missing values, or out-of-distribution samples to see how stable it is.
- **Comparative Benchmarks:** If relevant, run the model alongside current best practices or other algorithms to show its added value.

- **Iterative Tuning:** If initial tests reveal shortcomings, retrain or fine-tune the model and retest. Document those iterations.

Execution must be thorough enough to populate the credibility report. Sponsors should log everything meticulously, because later steps require justification of any decisions (see Step 6). The draft guidance emphasizes that *FDA engagement is important during execution* (www.dlapiper.com). This means sponsors may present interim results to FDA in meetings, get feedback on their plan, and adjust if necessary. For example, if FDA review staff thinks a planned validation dataset is too small, they might request a larger study.

From a project management view, Step 5 is often the longest and most resource-intensive phase, especially for complex models. It may involve statisticians, data scientists, and subject matter experts. It is critical that execution follows the plan; if it diverges (e.g. unforeseen issues force a changed test method), those changes must be documented as plan deviations in the next step.

Step 6: Document the Results

Once the plan is executed, Step 6 is to **document all findings in a credibility assessment report**. This report should compile the evidence, explaining how it was gathered and what it shows about the model's reliability. Main contents of the report include (www.dlapiper.com):

- **Description of Activities:** A record of what was done (e.g. "We tested the model on 200 independent subjects' data from Trial X").
- **Results Summary:** Tables/figures of performance metrics, error analyses, confusion matrices, ROC curves, etc.
- **Bias and Subgroup Analysis:** If performed, results of checks for consistent performance across demographic or disease subgroups.
- **Deviations from Plan:** Any differences from the original plan (Step 4) should be explained here. For example, "due to insufficient data, the test dataset was smaller than planned." Sponsors must justify any deviations and discuss how they affect confidence in the model.
- **Regulatory Considerations:** If discussions were had with FDA, summarize any agreements about data sufficiency or additional tests.
- **Limitations:** Note any known limitations of the validation. For instance, "the model has not been tested in patients over age 75." Transparency about limitations helps FDA understand applicability.

The credibility report is meant to provide *"information that establishes the credibility of the AI model for the COU"* (www.dlapiper.com). It should be clear and comprehensive: an FDA reviewer unfamiliar with the model should be able to conclude how well it works. Crucially, this is **not** a marketing document: it may contain negative findings if tests failed, but any such issues must be addressed. For example, if a model's accuracy dips to 80% vs the target 90%, the sponsor should note this and possibly explain how the model will still be managed (maybe requiring additional confirmatory tests in practice).

During early consultation, the draft recommends discussing with FDA whether, when, and where to submit the credibility assessment report. Depending on the engagement option, the report may be included in a regulatory submission or meeting package, or held and made available to FDA on request, including during an inspection (FDA draft, Step 6).

Step 7: Determine Adequacy of the Model for the COU

The final step is **drawing conclusions** on whether the model's performance and evidence meet the needs of the COU. In other words, *is the model credible enough to be used as intended?* This determination should weigh the evidence in the credibility report against the risk profile and the original question.

Possible outcomes of Step 7 include:

- **Adequate Credibility:** If the model met its performance objectives and no significant issues arose, the sponsor concludes the model is adequate. They may then proceed to rely on it in regulatory submissions or operational decisions, documenting this conclusion in their filings. They should also continue to monitor the model (passing data from actual use back into the model lifecycle plan).
- **Borderline/Insufficient:** If model credibility is not sufficiently established for the model risk, the draft describes five possible outcomes: (1) reduce model influence by incorporating additional types of evidence; (2) increase the rigor of credibility assessment activities or add development data; (3) establish appropriate risk controls; (4) change the modeling approach; or (5) consider the model output's credibility inadequate and reject or iteratively revise the context of use (FDA draft, Step 7). These options address inadequate credibility; the list does not offer deployment with monitoring alone as a fallback.

Whatever conclusion is reached, it should be justified by data. Sponsors should frame Step 7 findings in the context of risk: a high-risk model with marginal performance may not be acceptable, whereas a moderate-risk model with similar performance might be. The decision also sets the stage for regulatory communication: if there are uncertainties, sponsors need to explain how they'll manage them, either via mitigations (like extra patient monitoring) or follow-up studies.

Overall, the 7-step process is designed to be **transparent and iterative**. It parallels good practices in engineering ("verification and validation") and in biostatistics. It encourages scientific rigor: define the hypothesis (question), plan the experiment (credibility plan), run it, and then see if the hypothesis holds. The novelty is applying this rigor to AI models in regulated contexts. Several industry analyses confirm that if sponsors follow these steps, the FDA expects a predictable review process (www.goodwinlaw.com) (www.dlapiper.com). It is worth noting that this approach is *non-prescriptive* in the sense that FDA does not mandate specific thresholds or tests; rather, it leaves flexibility so that plans can be tailored. For example, it does not say "learning rate must be <0.1" or "use exactly 10,000 data points," since optimal methods vary by model and domain. Instead, it provides a **structured framework** so that sponsors and FDA can have a common language about AI credibility.

Contextual Considerations

Defining Credibility in AI Context

The concept of *credibility* is central to the guidance. FDA defines an AI model's credibility as "trust in the performance of an AI model for a particular context of use" (www.fda.gov). This echoes ideas from computational science (like the NASA credibility framework) and medical device V&V. In essence, credibility

means that the model does what it is expected to do, and does it consistently, for the specific question at hand.

Building credibility involves showing that the model is **valid** (it accurately solves the intended problem) and **reliable** (it does so repeatedly under the specified conditions) (www.bioprocessonline.com) (www.dlapiper.com). Sponsors can think of credibility as akin to analytical or clinical validation in other contexts. For example, an assay or lab test must be validated for accuracy and precision before being trusted; similarly, an AI model must be validated through evidence. FDA does not prescribe a specific "credibility score," but expects that the truth of the model's performance is demonstrated through data in the credibility report.

Several expert commentaries underline that credibility is not static. One article describes credibility as a continuum requiring *continuous oversight* with model updates and real-world monitoring (www.bioprocessonline.com). FDA's inclusion of a step on documenting results (and deviations) and its emphasis on life-cycle suggest that credibility must be maintained, not just achieved once. For example, after the AI model is used in manufacturing or postmarket, any significant retraining or drift would trigger a re-assessment or updating of the credibility documentation.

Thus, **expected practices** to ensure credibility include: robust cross-validation with separate data; transparency in algorithm design (to the extent possible); testing under varied conditions; monitoring outputs for anomalies; and updating models with new data responsibly. While the draft guidance does not mandate interpretability or explainability methods, the spirit of credibility implies that sponsors should at least understand their model's key drivers (e.g. using model explainability tools) to verify it aligns with scientific understanding.

AI Model Risk is Context-Specific

It bears repeating that *risk* is context-specific. The same AI algorithm might be used for very different ends. For example, a deep neural network segmenting medical images could be part of a device system (excluded from this guidance as it would be regulated as a device) or be used to quantify tumor changes for a drug trial primary endpoint (included here). Each use-case would have its own COU and risk profile.

Sponsors should categorize their AI use-cases similarly to earlier table: high-risk (e.g. life/safety decisions, final determinations) versus low-risk (e.g. supplementary analyses). DLA Piper notes that *final decisions without human intervention* are considered high risk (www.dlapiper.com). Another way to think of it: whenever patient welfare or product integrity could be compromised by model error, thorough scrutiny is needed. When an AI model is one part of a decision-making chain (with human oversight and fallback), the team can justify a lighter validation, but must still make sure the chain as a whole is safe.

AI model risk should also consider the maturity of the technology. Cutting-edge novel models may inherently carry more unknowns and thus be treated as higher risk. Conversely, a well-established statistical method (even if it qualifies as AI) might be seen as lower risk. That said, even a transparent model (like logistic regression) used in a high-stakes context requires careful validation; and a complex neural network used in a casual context could be overkill.

Relation to Other Regulations

This guidance does not exist in isolation. For AI-enabled **medical devices**, the FDA has separate guidance (finalizing a framework for AI/ML-driven software as a medical device) on how to handle software modifications and learning loops. The principles overlap (both emphasize risk, validation, and transparency) but the device guidance is tailored to software's role in diagnosis/treatment. Drug sponsors should be aware of any overlaps: if an AI model for drug use also relates to a medical device (e.g. a companion diagnostic), the corresponding device guidance may apply.

Importantly, the FDA clarifies in the draft that it is issuing *non-binding recommendations*. This means that legally, companies are not compelled to follow every suggestion or method. However, deviating from the guidance without justification is risky: since it represents FDA's thinking, ignoring it could trigger questions at review. If sponsors use alternative approaches (e.g. a different risk framework than "7-step"), they should clearly justify why their method is equally robust. The FDA's stated intent is to harmonize expectations, not to constrain innovation. Indeed, the agency solicits comments on whether more detailed guidance is needed for post-market AI use, indicating the process may evolve.

Data, Model Validation, and Evidence Considerations

A credible AI model depends on **data quality** and **rigorous validation**. Sponsors should ensure that training and testing datasets are appropriate and representative of the COU. For example, if an AI predicts outcomes for a certain patient population, the training data should cover that population's demographics and disease variations. If there are known subgroups (e.g. different age groups or genetic variants), the data should ideally include them, or the model's limitations in each subgroup should be documented.

FDA expects that sponsors will apply systematic verification and validation (V&V) methods to AI models, akin to other regulated tools. Verification means checking the model was implemented correctly (code review, unit tests, etc.), and validation means checking it yields clinically or scientifically correct output. While the guidance does not explicitly use the term "validation," Steps 4–6 are essentially a validation process: testing the model to confirm that it works as claimed.

Some specific evidence considerations:

- **Training/Tuning Data vs. Test Data:** Sponsors should clearly separate datasets. Data used to train the model (including any used for hyperparameter tuning) must be distinct from data used for evaluation. Overfitting is a known risk: a model might perform perfectly on training data but fail on new data. Thus, performance metrics should come from an unbiased test set. If proprietary or limited data means test sets are small, sponsors should discuss this limitation.
- **Real-World Data (RWD):** If the model uses RWD (electronic health records, insurance claims, etc.), the data's provenance and quality (coding accuracy, completeness) should be assessed. The guidance notes AI can analyze large real-world datasets (www.fda.gov), but sponsors must be mindful of biases in such data (e.g. undercoding of certain populations) and adjust accordingly.

- **Model Updates and Continual Learning:** Section IV.B addresses life cycle maintenance, including ongoing monitoring of model performance and evaluation of changes. In manufacturing, changes that may affect model performance are evaluated through the manufacturer's change management system within its pharmaceutical quality system. Depending on a change's extent and impact, credibility assessment activities may need to be repeated, including retraining and retesting. Changes affecting model performance should be reported to FDA in accordance with applicable regulatory requirements (FDA draft, Section IV.B).
- **Reproducibility and Audit Trails:** In a regulated environment, sponsors should maintain audit trails for AI development. This includes version control of code, records of parameter settings, and logs of when models were trained or retrained. This information helps if questions arise during review or inspection. The guidance expects transparency: in Step 4, the model description should enable an auditor to understand what was done.
- **Statistical Calibration:** Especially for predictive models, calibration (agreement between predicted probabilities and observed frequencies) is important. For example, if an AI model predicts a safety event with 10% probability, but the event actually occurs 30% of the time in test data, the model is miscalibrated. Calibration assessment is typically part of validation in clinical models. The draft explicitly asks sponsors to explain any calibration of the AI model under Section IV.A.4.a.iii, using a broader description than the probability-calibration example above (FDA draft, printed page 13).
- **Generalizability:** FDA recognizes that the real world can differ from controlled studies. Sponsors should think about how well the model will generalize to different settings. This might involve testing the model on external data (if available) or simulating broader conditions. For instance, if a model was trained on clinical trial data from one country, will it work equally well for another population? Any known limitations in generalizability should be disclosed.

In anticipation of these concerns, many commenters on AI in healthcare have stressed explainability and fairness. The draft guidance does not explicitly require explainable AI, but the notion of credibility implicitly demands that sponsors understand their models. Some sponsors may choose to apply explainable AI techniques (like feature importance or SHAP values) to illustrate that the model's drivers are logical. In regulated contexts, being able to explain a model's reasoning can increase trust, even if technically not mandated.

Statistical oversight also intersects: whenever AI models use patient data, FDA's Good Clinical Practice (GCP) and data privacy regulations still apply. For example, data used for AI training might need to be de-identified or meet HIPAA standards. These requirements are outside the draft's scope, but they should not be neglected when building an AI system.

Finally, sponsors should be aware that AI model failure modes can be subtle. Issues like data leakage (where training inadvertently includes future information), label errors (incorrect classification in training data), or subtle shifts in data distribution (e.g. a new diagnostic test with different calibration) can undermine performance. The credibility plan and report should demonstrate that such pitfalls were considered and mitigated.

Case Studies and Example Scenarios

To illustrate how the draft guidance applies, we consider several example use-cases drawing on published analyses and hypothetical scenarios. These examples span different stages of development and risk levels.

1. **High-Risk Clinical Decision AI (Hypothetical)**: A biotech develops an AI algorithm to classify patients with a novel oncology drug into “high-risk” or “low-risk” categories based on genetic markers and early lab results. This classification directly determines the treatment arm: high-risk patients get an intensive combination therapy, low-risk patients get standard therapy. This AI model is **in scope** because it impacts efficacy and safety decisions. (Step 1: QI = which arm to assign; Step 2: COU = stratify patients for risk-adapted therapy.) Because the AI’s decision could be life-altering, both influence and consequence are high (the AI has final say without human override). Model risk is therefore high. Under the guidance, the sponsor would need an extremely robust credibility plan: large, well-annotated training data from previous trials, multiple independent validation sets, and rigorous bias analysis. They might use cross-validation on past patient datasets and prospective validation in a pilot trial. Any failure (e.g. if the AI misclassifies a patient who then relapses) would be unacceptable. The sponsor would likely engage FDA early (perhaps at pre-IND) to agree on the plan (www.dlapiper.com). After execution, if the model beats pre-set benchmarks (say >95% accuracy on held-out test patients and balanced performance across subgroups), Step 6 documentation would detail these results. Step 7 would likely conclude adequacy, given the severity of risk was mitigated by strong performance. In practice, the sponsor may incorporate additional safeguards: for example, “all predictions must also be confirmed by a second method.” The developer might risk-reduce by not letting the AI fully automate the decision (thus reducing influence). Throughout, detailed documentation and early regulatory communication are key.

2. **Manufacturing Quality Control AI (FDA Hypothetical)**: FDA’s manufacturing example illustrates how release testing can reduce model influence. Measuring fill volume through release testing lowers model influence to low, while the decision consequence remains high. With those mitigations, the model risk for the context of use is medium (FDA draft, Step 3). FDA identifies its clinical-development and manufacturing examples as hypothetical; they do not extend beyond Step 3.

3. **Post-market Pharmacovigilance AI (Emerging Trend)**: Some pharmaceutical companies are exploring AI to automatically analyze large-scale safety data (e.g. electronic health records, social media reports) to detect adverse event patterns quicker than traditional methods. Suppose a company has an AI system that scans EHR databases and flags statistically significant spikes in certain side effects. This system could be used to generate alerts for new safety signals. Since patient safety is at stake, such an AI would be considered **in-scope**.

For example, if the AI catches a pattern suggesting a rare cardiac issue, regulators would want strong evidence the pattern is real (not an artifact of data). In Step 3, model influence is medium (the AI alerts but human investigators verify), and consequence is high (missed safety signals can harm patients). So overall risk is medium-high. The credibility plan might include back-testing the AI on known historical safety events (Step 4), showing it would have caught previous signals in retrospective data. It would also include a plan to

continue to validate with real-time data (Step 5). Step 6 documentation could present metrics like sensitivity in detecting known signals. Step 7 would involve agreeing on how to act on AI alerts (will the company submit them as signals to FDA, or is this for internal monitoring?).

This scenario highlights an important nuance: AI in **postmarketing** is a developing area. The FDA guidance explicitly asks for comment on whether more detail is needed for AI in pharmacovigilance (www.goodwinlaw.com). That suggests regulators see this as an evolving field. For now, sponsors should treat it as in-scope but can expect FDA comments. Real-world examples are limited, but one can imagine regulators expecting high transparency: e.g., how the AI handles reporting biases, how cases are adjudicated after flagging, etc.

4. **Medical Writing Support AI (Excluded Use)**: A company uses a generative language model to draft sections of their Clinical Study Report and Investigator's Brochure. Scientists provide bullet-point data, and the AI produces narrative text. Humans then edit the text. This use falls outside the draft's scope only when the operational-efficiency use does not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). Human editing alone does not replace those scope conditions. The exclusion concerns this draft and does not establish a general exemption from applicable regulatory requirements.

FDA's scope language includes drafting or writing a regulatory submission among operational-efficiency uses, subject to the conditions on patient safety, drug quality, and study-result reliability. Sponsors uncertain whether their AI use meets the draft's scope conditions are encouraged to engage FDA early (FDA draft, Section II).

5. **Drug Discovery AI (Excluded vs. Included Nuance)**: Consider an AI model that proposes novel molecular structures with predicted drug-like properties. This model is used **before any preclinical studies**. By the guidance's language, this is explicitly excluded (regulations.justia.com). The rationale is that at the discovery stage, outputs are not directly evidence for safety or efficacy; one still does lab tests afterwards.

However, if an AI in discovery also provides output that *later* feeds into a regulated decision, nuances arise. For example, if the AI also predicts a toxic metabolite that will influence later safety testing, one could argue that component of the AI's output has a regulatory impact. The FDA's wording implies that basic discovery is out-of-scope, but sponsors should be cautious: any AI output that eventually affects the safety profile determination (like predicting toxicology study results) should be included under the 7-step process for that context. In practice, most discovery-stage AI falls outside the scope, but once the project crosses into preclinical candidate selection, any AI-derived evidence should be handled per the guidance.

As an illustrative nuance, the Foley/NatLawReview commentary noted that "the guidance emphasizes that it is limited to AI models (including for drug discovery) that impact patient safety, drug quality, or reliability of results" (natlawreview.com). This suggests that *some* aspects of discovery-driven AI could be in scope if they directly address safety/quality questions. Still, the safe interpretation is: AI suggestions for target chemistry (pure research use) are outside; model analyses of preclinical data (regulatory use) are inside.

These examples illustrate how context of use determines the draft's scope. The draft addresses AI models producing information or data to support regulatory decisions about drug safety, effectiveness, or quality. It does not address drug discovery or operational-efficiency uses that do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). Its 7-step framework is proposed guidance, not a mandatory procedure; alternative approaches may be used if they satisfy applicable statutes and regulations.

Perspectives and Implications

The draft guidance has stirred discussion among various stakeholders. Here we summarize the perspectives of industry sponsors, legal advisors, and regulators, and consider the broader implications.

Industry Sponsors and Developers

For pharmaceutical companies and biotech firms, the guidance provides much-needed clarity. Previously, sponsors had to guess what documentation FDA might require when using AI. Now they have a formal roadmap. Many life-sciences companies are excited about AI's potential but were held back by regulatory uncertainty. An industry executive noted that regulatory clarity was among the "top three barriers" to implementing AI in drug development (www.statnews.com). By codifying expectations, the FDA lowers that barrier.

Companies will need to adjust their development practices. The draft offers nonbinding recommendations for assessing AI model credibility; it does not itself impose new development-process requirements (FDA guidance status). Biostatistics and engineering teams will likely take on more prominent roles in proof-of-concept studies. Development plans should incorporate FDA's 7 steps: e.g., IND planning should specify if an AI model will be used and include a credibility plan outline. For companies already using AI, retrospective documentation of credibility activities may need to be compiled. Smaller companies and startups might face challenges due to limited resources for such thorough validation, making it harder to include AI early. However, many sponsors may view this as worthwhile to gain regulatory confidence.

On the technical side, some sponsors fear that the guidance's non-prescriptive nature could be interpreted inconsistently. For example, what constitutes "adequate" validation for a given model? The FDA says plans must be "commensurate with risk," but does not give numeric criteria. As a result, companies may run mock audits internally or seek pre-IND meetings to gauge FDA staff expectations. Over time, as case precedents (or FDA-personnel guidance) emerge, standards will likely crystallize. Some sponsors may greatly exceed minimal requirements to avoid any question, adopting best practices like third-party audits of model development, extensive bias assessments, or even industry consensus standards for AI (if those develop).

The draft's exclusion for operational efficiencies, including drafting or writing a regulatory submission, applies when those uses do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results (FDA draft, Section II). This limited scope exclusion does not establish that administrative AI is outside FDA authority or provide immunity from applicable regulatory requirements.

The draft does not create new legally enforceable obligations for startups or technology vendors (FDA draft). Its recommendations should be distinguished from applicable statutory or regulatory requirements. For example, an AI company selling a patient stratification tool may need to share training methodology and validation data with the sponsor. This raises questions about proprietary algorithms and data privacy. Companies will need to balance transparency with IP protection. It is probable that service agreements will include clauses for "supporting regulatory review" of AI performance.

Regulatory and Policy Perspective

From FDA's viewpoint, this guidance embodies the agency's intention to be agile and risk-based. Commissioner Califf and others have emphasized that they want to "promote innovation and ensure the agency's robust scientific standards are met" (www.fda.gov). The 7-step framework reflects a collaboration between multiple FDA offices, as noted in the Federal Register preamble (regulations.justia.com). By engaging stakeholders early (even while the draft was being written, presumably), FDA aimed to produce guidance that is implementable rather than idealistic.

One notable aspect is the invitation for public comment and the openness to feedback on specific issues (e.g., alignment with industry experience, post-market AI use) (www.goodwinlaw.com). This suggests the guidance may evolve. FDA explicitly asked if the 7-step approach aligns with sponsors' real-world experience, and whether further guidance is needed for life-cycle maintenance or pharmacovigilance. The comment period (through April 2025) could result in changes. For instance, sponsors may request numeric thresholds or examples of adequate evidence, while patient advocacy groups may push for stronger emphasis on fairness and transparency.

Internationally, the FDA guidance may influence other regulators. Firms often approach global submissions with an eye toward regulatory harmonization. While the FDA and EMA documents are not identical (EMA took a broader principles approach in its reflection paper (www.ema.europa.eu)), both emphasize risk and evaluation. Sponsors developing global strategies should ensure their credibility frameworks can satisfy multiple agencies. For example, if testing in Europe as well, companies may cite FDA's guidance for their US discussions and EMA's paper for EU bodies.

Expert and Industry Commentary

Early analyses of the FDA draft from law firms and consultancies have generally noted both its innovativeness and challenges. The American Bar Association's Health Law section observed the "potential seven-step, risk-based framework" and noted the carve-outs for discovery and efficiencies (www.americanbar.org). They viewed the draft as a cautious step to balance innovation and protection. The Foley & Lardner commentary highlighted that sponsors may have to prepare detailed AI documentation (data and governance) in submissions (www.foley.com).

A DLA Piper white paper spelled out key takeaways and explicitly quoted FDA's carve-outs (www.dlapiper.com). It also enumerated some practical strategies: for instance, at Step 6 they note that during early FDA consultations, sponsors should clarify if and how to submit the report.

Regulatory affairs experts generally agree that the guidance is a “good start” but is still just a draft. Some commentaries (e.g. from Bioprocess Online (www.bioprocessonline.com)) stress that AI model maintenance and updating deserve more emphasis. Others suggest that terminologies like “internal workflows” should be defined clearly, so companies don’t misinterpret the liminal cases. The draft’s focus on context-specific risk is welcomed as science-based, but some have called for illustrative examples or case studies to be included for clarity.

Impact on Patients and Public Health

Ultimately, the goal of regulating AI in drug development is to protect patients. If AI tools churn out flawed analyses, patients could be harmed by unsafe drugs or denied beneficial therapies. The credibility framework aims to minimize this risk by raising the evidentiary bar. The draft offers nonbinding recommendations and is not for implementation (FDA guidance status).

On the flip side, if properly implemented, this guidance could accelerate safe innovation. Patients could benefit sooner from new treatments because AI streamlines research (as long as there is confidence in the results). For example, AI could help identify rare responders to a therapy, enabling smaller, faster trials. The draft guidance encourages such uses by providing a clear pathway: sponsors know that if they follow the 7 steps, the regulatory review of their AI-augmented data should proceed smoothly.

However, there is a potential risk that the added burden might slow down some AI applications in the short term. Especially for smaller companies, conducting all the validation activities might delay product development. Over time, though, as best practices emerge and AI becomes more standardized, these hurdles may lessen. The draft proposes recommendations rather than codifying new requirements (FDA guidance status).

An important patient-centric implication is that AI use in submissions will be documented and scrutinized. In cases where an AI model plays a major role, patients (and healthcare providers) can be assured that there is disclosure and analysis behind it. For example, if a drug labelling says efficacy was determined using an AI model “validated per FDA guidance,” this transparency is good for trust. Patient advocacy groups may push for consumer-friendly explanations of how AI was used in regulatory decisions. The framework sets the stage for this kind of accountability.

Future Directions

Several avenues could shape the future of AI regulation in drug development:

- **Finalization of Guidance:** The FDA will likely issue a final guidance after reviewing comments from industry and other stakeholders. Comments may refine the document’s language, perhaps adding clarifications, examples, or even tightening definitions (e.g., explicitly defining “operational efficiency”). The final guidance might also address some of the open questions – for instance, how to handle AI in post-market safety monitoring, or whether to expect periodic updates to the credibility plan.

- **Emerging AI Technologies:** Since January 2025, new AI paradigms (like advanced large language models, multimodal systems, etc.) have emerged. Generative AI (ChatGPT, DALL-E, etc.) is becoming more capable. The draft excludes operational-efficiency uses such as drafting or writing a regulatory submission when they do not impact patient safety, drug quality, or the reliability of nonclinical or clinical study results; it does not establish an unconditional exclusion for content creation (FDA draft, Section II). Regulatory policy will likely evolve to address novel technologies. For example, if an LLM were used as a decision support in evaluating preclinical data, would that be in scope? Over time, FDA guidance or Q&A may interpret the draft's principles in light of new tech.
- **Standards and Best Practices:** Expect the development of industry standards for AI validation. Groups like the International Council for Harmonisation (ICH) or ISO may work on guidance for AI models. If uniform standards emerge (e.g. a standard set of bias metrics or validation criteria for clinical AI), FDA and other regulators might adopt them. Being early with a credible framework could position FDA to influence such standards globally.
- **Integration with Digital Health Initiatives:** The line between drugs and digital therapeutics may blur. If a drug's effectiveness is partly mediated by an AI app (e.g. an AI that personalizes a diet or exercise regimen complementary to a drug), regulators will need to consider AI software in the regulatory mix. The current guidance might not fully cover combined drug-digital products, so that could be an area for future rules.
- **Post-market Surveillance of AI:** As AI tools begin to be used, regulators will want to monitor their real-world performance. The guidance hints at life-cycle maintenance, but doesn't detail post-market AI oversight. In the future, FDA may request periodic updates on deployed models, similar to post-market reporting for drugs (like annual reports), especially if the AI is adaptive. There is also the question of how FDA might audit AI systems (e.g. conducting software inspections or requiring submission of logs).
- **Global Regulatory Harmonization:** Harmonization has already advanced: EMA and FDA jointly identified ten principles for good AI practice in the medicines lifecycle in January 2026 (EMA announcement; joint principles). Separately, FDA lists ICH M15, *General Principles for Model-Informed Drug Development*, as final guidance dated June 2026; it provides a broader framework for MIDD evidence (FDA M15; FDA MIDD program). Future AI-specific guidance may build on these existing initiatives.
- **Innovation Incentives:** Recognizing AI's potential value, regulators and policymakers might introduce incentives for responsible AI. For example, there could be programs that provide "AI validation grants" for smaller companies, or expedited review pathways for therapies using rigorously validated AI. Alternatively, if AI proves essential to modern drug research, FDA might integrate AI assessment into existing expedited programs (like Breakthrough Therapy designation) to streamline AI oversight.

In sum, the January 2025 draft guidance is a landmark but not an endpoint. Its publication signals an ongoing evolution. The FDA and industry will continue to iterate on how best to benefit from AI while guarding patients. As one commentator noted, this guidance incorporates "comprehensive recommendations for the design, development, documentation, and maintenance of these AI models" (www.goodwinlaw.com), foreshadowing that the principles laid out today will form the backbone of AI regulation in life sciences for years to come.

Conclusion

The FDA's January 2025 draft guidance on AI in drug development marks a pivotal moment in regulatory science. It is the first formal blueprint for how the FDA expects sponsors to handle AI tools whose outputs inform safety, efficacy, or quality decisions. By introducing a **risk-based, 7-step credibility framework**, the guidance provides a clear roadmap: define the question and context, assess risk, plan and conduct validation, and document outcomes. Importantly, the guidance also explicitly **limits its scope** to ensure that only AI with potential patient/quality impact is covered, while exempting purely internal or discovery uses (regulations.justia.com) (www.dlapiper.com).

Throughout this report we have examined the details of the guidance and surrounding analysis. We have highlighted how it builds on regulatory precedents, how companies can prepare for its proposed recommendations (FDA guidance status), and what it means for future innovation. We showed that sponsors must distinguish between AI for *evidence generation* (in scope) and AI for *operational efficiency* (out of scope), and we have described the evidence and documentation needed to satisfy FDA that an AI model is "credible" for its purpose. With extensive citations, we underscored the significance of each element of the guidance.

Case examples illustrate that the guidance's application can be intuitive: high-risk AI in trials needs robust validation, while advisory AI in manufacturing needs just enough evidence. We also noted areas needing attention, such as AI in pharmacovigilance, where FDA is actively seeking input.

Looking forward, this guidance is likely to shape industry practices for the foreseeable future. It invites sponsors to proactively integrate AI evaluation into their development pipelines, offers regulators a consistent way to review AI-backed data, and aims to reassure all stakeholders that innovation does not come at the cost of safety or quality. As AI technology continues to advance, both FDA and the biopharma industry will watch closely how this framework is received, refined through public comment, and implemented in real-world submissions.

In conclusion, the FDA's draft guidance is a comprehensive attempt to bridge the gap between cutting-edge AI possibilities and well-established regulatory standards. The draft provides a risk-based framework that may be used to establish and evaluate an AI model's credibility for a particular context of use (FDA draft). Sponsors who align with this framework will help ensure that AI-driven drug development remains safe, effective, and scientifically sound.

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/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions
2. https://www.dlapiper.com/en-sa/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai
3. https://regulations.justia.com/regulations/fedreg/2025/01/07/2024-31542.html
4. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
5. https://www.fda.gov/media/184830/download

6. https://www.statnews.com/2025/01/06/artificial-intelligence-use-drug-development-fda-draft-guidance
7. https://www.foley.com/de/insights/publications/2025/01/ai-drug-development-fda-releases-draft-guidance
8. https://www.goodwinlaw.com/en/insights/publications/2025/01/alerts-lifesciences-aiml-fda-publishes-its-first-draft-guidance
9. https://www.ema.europa.eu/en/use-artificial-intelligence-ai-medicinal-product-lifecycle-scientific-guideline
10. https://www.americanbar.org/groups/health_law/news/2025/1/fda-published-draft-guidance-ai-use-drug-biological-product-submissions
11. https://talglaw.com/artificial-intelligence-for-drug-biological-products-credibility-of-ai-models-for-context-of-use-the-model-risk-paradigm
12. https://www.bioprocessonline.com/doc/deciphering-fda-s-step-framework-for-ai-driven-decision-making-0001
13. https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device
14. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfstandards/detail.cfm?standard__identification_no=38534
15. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/m15-general-principles-model-informed-drug-development
16. https://www.fda.gov/medical-devices/digital-health-center-excellence/step-6-software-function-intended-provide-clinical-decision-support
17. https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan
18. https://natlawreview.com/article/ai-drug-development-fda-releases-draft-guidance
19. https://www.ema.europa.eu/en/news/ema-fda-set-common-principles-ai-medicine-development-0
20. https://www.ema.europa.eu/system/files/documents/other/2026-01-13_ai-guiding-principles_ema-fda_en.pdf
21. https://www.fda.gov/drugs/development-resources/model-informed-drug-development-paired-meeting-program

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

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.