Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Tala Fakhouri, 301-837-7407; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Digital Health Center of Excellence, digitalhealth@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs
and/or
Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances
and/or
Office of Policy
Center for Devices and Radiological Health
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 66, Room 5431
Silver Spring, MD 20993-0002
Email: CDRH-Guidance@fda.hhs.gov
https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/guidance-documents-medical-devicesand-radiation-emitting-products

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

TABLE OF CONTENTS

Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry[1] and Other Interested Parties

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

This guidance provides recommendations to sponsors[2] and other interested parties[3] on the use of artificial intelligence (AI) to produce information or data intended to support regulatory decision-making[4] regarding safety, effectiveness, or quality for drugs.[5,6] Specifically, this guidance

[1] This guidance has been prepared by the Center for Drug Evaluation and Research in collaboration with the Center for Biologics Evaluation and Research, the Center for Devices and Radiological Health, the Center for Veterinary Medicine, the Oncology Center of Excellence, the Office of Inspections and Investigations, and the Office of Combination Products in the Office of the Commissioner at the Food and Drug Administration (FDA).

[2] Depending on the stage of the drug product life cycle, FDA may refer to a person or entity as a *sponsor, a requestor,* or an *applicant.* For example, a *sponsor* may refer to a person or an entity that takes responsibility for and initiates a clinical investigation. The terms *requestor* and *sponsor* are used in various contexts for over-the-counter monograph drugs. An *applicant* may refer to the person or entity that files a marketing application and/or assumes responsibility for the marketing of a human drug, animal drug, or biological product. Because this guidance covers the drug product life cycle, including premarket and postmarketing activities, this guidance uses the single term *sponsor* to cover sponsors, requestors, and applicants, as applicable.

[3] For the purposes of this guidance, an *interested party* means any person or organization that may be interested in the use of AI in drug and biological product development. This includes, for example, manufacturers (i.e., a person or entity that manufactures, processes, packs, or holds a drug) that are otherwise not sponsors.

[4] For the purposes of this guidance, *regulatory decision-making* refers to regulatory determinations made by FDA (e.g., with respect to an application or supplement) and actions taken by sponsors and other interested parties in conformance with FDA's regulatory authority (e.g., current good manufacturing practices (CGMPs), postmarketing requirements, investigational new drug applications (INDs).

[5] For the purposes of this guidance, the term *drug*, as defined in section 201(g) of the Federal Food, Drug, and Cosmetic Act (FD&C Act), refers to human and animal drugs and human biological products (as defined in section 351(i) of the Public Health Service Act), other than biological products that also meet the definition of a device under section 201(h)(1) of the FD&C Act, unless otherwise specified. It also refers to a drug or biological product constituent part (21 CFR 4.2) of a combination product (21 CFR 3.2).

[6] The recommendations in this guidance focus on the use of AI to produce data or information to support regulatory decision-making for drugs or combination products that include a drug. The recommendations also may be relevant across all medical products, including to support regulatory decision-making for medical devices intended to be used

provides a risk-based credibility assessment framework that may be used for establishing and evaluating the credibility of an AI model[7] for a particular context of use (COU). For the purposes of this guidance, credibility refers to trust, established through the collection of credibility evidence, in the performance of an AI model for a particular COU. Credibility evidence is any evidence that could support the credibility of an AI model output for a specific COU. The COU defines the specific role and scope of the AI model used to address a question of interest. This guidance does not endorse the use of any specific AI approach or technique.

As used in this guidance, AI refers to a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.[8] AI systems (1) use machine- and human-based inputs to perceive real and virtual environments, (2) abstract such perceptions into models through analysis in an automated manner, and (3) use model inference to formulate options for information or action.[9] A subset of AI that is commonly used in the drug product life cycle[10] is machine learning (ML). ML refers to a set of techniques that can be used to train AI algorithms to improve performance at a task based on data.[11] Although ML is currently the most utilized AI modeling technique in the drug product life cycle, this guidance focuses on AI models more broadly.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of

---

with drugs. The term *device* refers to a device as defined in section 201(h)(1) of the FD&C Act (21 U.S.C. 321(h)(1)). For devices, FDA recommends that sponsors refer to device-specific guidances using CDRH's guidance search web page Guidance Documents (Medical Devices and Radiation-Emitting Products) at https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/guidance-documents-medical-devices-and-radiation-emitting-products.

[7] Depending on the intended use (see 21 CFR 801.4) of an AI model, the AI model may meet the definition of a device under section 201(h)(1) of the FD&C Act. How to determine whether an AI model meets the definition of a device is outside the scope of this guidance. For further information about FDA digital health regulatory policies, see FDA's web page Digital Health Policy Navigator at https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-policy-navigator and FDA's web page on Guidances with Digital Health Content, https://www.fda.gov/medical-devices/digital-health-center-excellence/guidances-digital-health-content.

[8] See Executive Order 14110 of October 30, 2023; Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, sec. 3(b) (citing to definition of AI at 15 U.S.C. 9401(3)); https://www.federalregister.gov/d/2023-24283.

[9] Ibid.

[10] For the purposes of this guidance, the term *drug product life cycle* includes nonclinical, clinical, postmarketing, and manufacturing phases. While the drug product life cycle generally also includes drug discovery, the use of AI for the purposes of drug discovery is not in the scope of this guidance and therefore is not included in our use of the term drug product life cycle.

[11] See Executive Order 14110 of October 30, 2023; Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, sec. 3(t) (definition of ML); https://www.federalregister.gov/d/2023-24283.

the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. SCOPE

This guidance discusses the use of AI models in the drug product life cycle, where the specific use of the AI model is to produce information or data to support regulatory decision-making regarding safety, effectiveness, or quality for drugs.

This guidance does not address the use of AI models (1) in drug discovery or (2) when used for operational efficiencies (e.g., internal workflows, resource allocation, drafting/writing a regulatory submission) that do *not* impact patient safety, drug quality, or the reliability of results from a nonclinical or clinical study. We encourage sponsors to engage with FDA early if they are uncertain about whether or not their use of AI is within the scope of this guidance.

The risk-based credibility assessment framework[12,13] described in this guidance is intended to help sponsors and other interested parties plan, gather, organize, and document information to establish the credibility of AI model outputs when the model is used to produce information or data intended to support regulatory decision-making. As described in this guidance, the activities (e.g., the level of oversight by FDA, the sponsor, or other parties responsible for the relevant information or data, the stringency of the credibility assessments and the performance acceptance criteria, the risk mitigation strategy, and the type and extent of documentation and detail associated with AI use) that may be used to establish credibility of AI model outputs should be commensurate with the AI model risk and tailored to the specific COU.

This guidance also describes different options by which sponsors and other interested parties may engage with the Agency on issues related to AI model use, depending on the COU and the specific development program.

[12] FDA applies benefit-risk principles when assessing the safety, effectiveness, and quality of a drug. For illustrative examples highlighting benefit-risk considerations, see (1) the guidance for industry *Benefit-Risk Assessment for New Drug and Biological Products* (October 2023), (2) the draft guidance for industry *Benefit-Risk Considerations for Product Quality Assessments* (May 2022) (when final, this guidance will represent FDA's current thinking on this topic), (3) the International Council for Harmonisation (ICH) guidance for industry *M4E(R2): The Common Technical Document (CTD)—Efficacy* (July 2017), (4) the ICH guidance for industry *Q9(R1) Quality Risk Management* (May 2023), and (5) the ICH guidance for industry *Q10 Pharmaceutical Quality System* (April 2009).

[13] The high-level key concepts and principles of the risk-based credibility assessment framework described in this guidance (sections IV.A.1 through A.3) were informed by an FDA-recognized consensus standard for medical devices titled "American Society of Mechanical Engineers Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Device" (ASME V&V40). While the ASME V&V40 was developed specifically for physics-based models for medical device applications, the high-level key concepts related to defining the question of interest, COU, and assessment of model risk, which are outlined in sections 2, 3, and 4 of the ASME V&V40 standard, are used in this guidance's risk-based credibility assessment framework for the use of AI models to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs.

III. BACKGROUND

In recent years, the use of AI in the drug product life cycle has increased. Continuous advancements in AI hold the potential to accelerate the development of safe and effective drugs and enhance patient care. Concurrent with these technological advancements, the use of AI in regulatory submissions to FDA has also increased for some uses.[14] Some examples[15] of AI uses for producing information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs include, but are not limited to, (1) reducing the number of animal-based pharmacokinetic, pharmacodynamic, and toxicologic studies; (2) using predictive modeling for clinical pharmacokinetics and/or exposure-response analyses; (3) integrating data from various sources (e.g., natural history, clinical studies, genetic databases, clinical trials, social media, registries) to improve understanding of disease presentations, heterogeneity, predictors of progression, recognition of disease subtypes; (4) processing and analyzing large sets of data (e.g., data from real-world data sources or data from digital health technologies) for the development of clinical trial endpoints or assessment of outcomes; (5) identifying, evaluating, and processing for reporting postmarketing adverse drug experience information; and (6) facilitating the selection of manufacturing conditions.

However, AI use presents some unique challenges. First, the variability in the quality, size, and representativeness of datasets for training AI models[16] may introduce bias and raise questions about the reliability of AI-driven results. As such, data used to develop AI models should be *fit for use*,[17] which means the data should be both relevant (e.g., includes key data elements and sufficient numbers of representative participants[18] or sufficient data that is representative of the

[14] See, e.g., Liu, Q, R Huang, J Hsieh, et al., 2023, Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021, Clin Pharmacol Ther, 113(4):771–774, doi:10.1002/cpt.2668.

[15] For more information on using AI and ML in the development of drug and biological products, see https://www.fda.gov/media/167973/download.

[16] For the purposes of this guidance, *training data* are data used in procedures and training algorithms to build an AI model, including to define model weights, connections, and components. These data typically should be representative of the target patient population or the manufacturing process or operation, as applicable. For further information regarding training data, see the guidance for industry and FDA staff *Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions* (December 2024) and the draft guidance to industry and FDA staff *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations* (January 2025). When final, this guidance will represent FDA's current thinking on this topic. We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/regulatory-information/search-fda-guidance-documents.

[17] The terms *fit for use* and *fit for purpose* are sometimes used interchangeably.

[18] Human subjects protections are out of the scope of this guidance but should be considered when developing or deploying AI modeling in the drug product life cycle, as applicable. For additional information, see FDA's web page Regulations: Good Clinical Practice and Clinical Trials at https://www.fda.gov/science-research/clinical-trials-and-human-subject-protection/regulations-good-clinical-practice-and-clinical-trials

manufacturing process or operation) and reliable (i.e., accurate, complete, and traceable).[19] Second, because of the complex computational and statistical methodology underpinning these models, understanding how AI models are developed and how they arrive at their conclusions may be difficult and necessitate methodological transparency (e.g., detailing in the regulatory submission the methods and processes used to develop a particular AI model). Third, uncertainty of the accuracy in the deployed models' output may be difficult to interpret, explain, or quantify. Finally, another challenge with some AI models is the potential for the model's performance to change over time or across deployment environments when new data inputs are introduced and these inputs differ from the data on which the model was trained (i.e., data drift) requiring life cycle maintenance of these models.

IV. CONSIDERATIONS FOR AI USE IN THE DRUG PRODUCT LIFE CYCLE

Section IV.A outlines the proposed risk-based credibility assessment framework for AI use in the drug product life cycle. Section IV.B discusses the importance of life cycle maintenance of the credibility of AI model outputs in certain contexts of use. Section IV.C describes different options by which sponsors, and other interested parties, may engage with the Agency on issues related to AI model development.

A. A Risk-Based Credibility Assessment Framework

Among various computational models used in the drug product life cycle, this guidance focuses on the use of AI models to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs.

The risk-based credibility assessment framework described here consists of the following 7-step process to establish and assess the credibility of an AI model output for a specific COU based on model risk:

- Step 1: Define the question of interest that will be addressed by the AI model (see section IV.A.1 for details).
- Step 2: Define the COU for the AI model (see section IV.A.2 for details).
- Step 3: Assess the AI model risk (see section IV.A.3 for details).
- Step 4: Develop a plan to establish the credibility of AI model output within the COU (see section IV.A.4 for details).
- Step 5: Execute the plan (see section IV.A.5 for details).

[19] For further information, see the guidance for industry *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products* (July 2024).

- Step 6: Document the results of the credibility assessment plan and discuss deviations from the plan (see section IV.A.6 for details).

- Step 7: Determine the adequacy of the AI model for the COU (see section IV.A.7 for details).

For steps 1 through 3, two examples will be used to illustrate the process of describing the question of interest, defining the COU, and demonstrating how model risk might be assessed. One example involves AI use in clinical development and the other involves AI use in manufacturing. These two hypothetical examples do not extend beyond step 3 because the credibility assessment activities listed in step 4 are intended to provide a general list of activities that should be considered when establishing the credibility of AI model outputs. The appropriate credibility assessment activities may vary depending on the nuances of a specific development program that cannot be captured in the hypothetical examples provided. Additionally, steps 5 through 7 relate to step 4, as they are intended recommendations to execute, document, and assess the credibility assessment activities of step 4. As such, hypothetical examples illustrate the concepts described in steps 1 through 3 only.

1. Step 1: Define the Question of Interest

Step 1 in the framework is to define the question of interest. The question of interest should describe the specific question, decision, or concern being addressed by the AI model.

As an example of defining the question of interest in clinical development, Drug A is under development and is associated with a life-threatening drug-related adverse reaction.[20] In previous trials for Drug A, all participants went through 24-hour inpatient monitoring after dosing due to concerns about this adverse reaction. However, data from these previous trials showed that some participants were at low risk for this adverse reaction. In a new study, the sponsor is exploring a strategy to use an AI model to stratify patients for 24-hour inpatient monitoring based on their risk for experiencing this adverse reaction. In the sponsor's proposal, participants with low risk for the adverse reaction will be sent home for outpatient monitoring after dosing. For this example, the question of interest would be "Which participants can be considered low risk and do not need inpatient monitoring after dosing?"

As an example of defining the question of interest in commercial manufacturing, Drug B is a parenteral injectable dispensed in a multidose vial. The volume is a critical quality attribute for the release of vials of Drug B. A manufacturer is proposing to implement an AI-based visual analysis system to perform 100% automated assessment of the fill level in the vials, to enhance

[20] The use of AI must comply with all applicable regulatory requirements. This includes, for example, in clinical development, section 505 of the FD&C Act, and 21 CFR parts 50, 56, and 312.

the performance of the visual analysis system and identify deviations.[21] For this example, the question of interest would be "Do vials of Drug B meet established fill volume specifications?"

A variety of evidentiary sources may be used to answer the question of interest. For example, evidence generated from, but not limited to, in vitro testing, in vivo animal testing, clinical trials, or manufacturing process validation studies may be used *in conjunction* with evidence generated from the AI model to address any specific question of interest. These different evidentiary sources should be stated when describing the AI model's COU in step 2 and are relevant when determining *model influence* as assessed in step 3. Sponsors should engage with FDA early if they are uncertain about their evidentiary sources.

2. Step 2: Define the Context of Use for the AI Model

Step 2 in the framework is to define the COU for the AI model. The COU defines the specific role and scope of the AI model used to address a question of interest. The description of the COU should describe in detail what will be modeled and how model outputs will be used. The COU should also include a statement on whether other information (e.g., animal or clinical studies) will be used in conjunction with the model output to answer the question of interest.

For example, to answer the question of interest in the clinical development example discussed in section IV.A.1 ("Which participants can be considered low risk and do not need inpatient monitoring after dosing?"), a sponsor is proposing to use an AI model to predict a participant's risk for the drug-related adverse reaction to Drug A based on baseline characteristics and lab values. Specifically, the output from the AI model will be used to stratify participants into low- versus high-risk groups for the potentially life-threatening adverse reaction to Drug A (the AI model's role). In this context, the sponsor is proposing that *only* the AI model will be used to determine whether the participant is considered low risk and whether they will need inpatient or outpatient monitoring after dosing (the AI model's scope). This would be considered the COU of the AI model for this example.

For the manufacturing example mentioned previously in section IV.A.1 (to answer the question of interest "Do vials of Drug B meet established fill volume specifications?"), an AI-based model will be used to analyze data obtained from visual images of the vials to determine if a deviation in volume has occurred (the AI model's role). However, as part of release testing, independent verification of the fill volume is performed on a representative sample for each batch. Therefore, the AI-based model will *not* be the sole determinant for the release of product (the AI model's scope). This is the COU of the AI model for this example.

[21] The use of AI in manufacturing (e.g., production and process controls) must be implemented in accordance with current good manufacturing practice (see section 501(a)(2)(B) of the FD&C Act and 21 CFR part 211). For example, with regard to finished drug products, the responsibilities of the quality control unit described in 21 CFR 211.22 and 211.68 are applicable. The quality control unit is ultimately responsible for ensuring the overall quality of the final drug product (see 21 CFR 210.3).

3. Step 3: Assess the AI Model Risk

Step 3 in the framework is to assess the model risk. Model risk is a combination of two factors: (a) model influence, which is the contribution of the evidence derived from the AI model relative to other contributing evidence used to inform the question of interest and (b) decision consequence, which describes the significance of an adverse outcome resulting from an incorrect decision concerning the question of interest.[22] Model risk is illustrated in Figure 1.

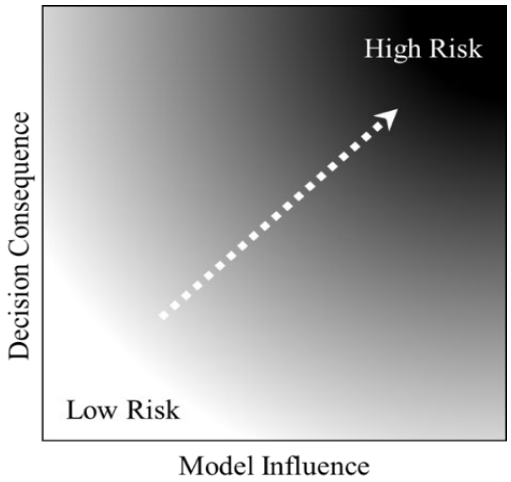

Figure 1. Model risk matrix. The model risk moves from low to high as decision consequence or model influence increases. The ratings for decision consequence and model influence are independently determined.

Model risk is the possibility that the AI model output may lead to an incorrect decision that could result in an adverse outcome, and not risk intrinsic to the model.[23] Assessing model risk involves subject matter expertise and judgment among sponsors and interested parties and FDA.

This model risk matrix can be applied to the clinical development example described in section IV.A.1 to address the question of interest "Which participants can be considered low risk and do not need inpatient monitoring after dosing?" In this example, model influence would likely be estimated to be high because the AI model will be the *sole* determinant of which type of patient monitoring a participant undergoes. The decision consequence is also high because if a participant who requires inpatient monitoring is placed into the outpatient monitoring category, that participant could have a potentially life-threatening adverse reaction in a setting where the participant may not receive proper treatment. Given that model influence is deemed high for this question of interest and decision consequence is also deemed high, the model risk for this COU is high.

[22] The decision consequence is the significance of an adverse outcome resulting from an incorrect decision concerning the question of interest. Decision consequence is the potential outcome of the overall decision that is made by answering the question of interest, outside of the scope of the AI model and irrespective of how modeling is used. That is, decision consequence should consider the question of interest, but should not consider the COU of the model. Additionally, when assessing the decision consequence, FDA recommends that sponsors consider both the potential severity of adverse outcome and the probability that the adverse outcome would occur. In some risk management tools, the ability to detect the harm (detectability) also factors into the estimation of risk. For more information, see the guidance for industry and FDA staff *Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions* (November 2023).

[23] Other types of risk, such as cybersecurity risk, are out of scope of this guidance but should be considered when deploying AI modeling in the drug product life cycle.

For the commercial manufacturing example described in section IV.A.1, deviations in the volume of vials containing Drug B could result in a number of issues. For example, the release of units that do not meet quality standards could potentially lead to medication errors due to either an inability to withdraw labeled content or pooling of vials to obtain a single dose (if not identified in labeling).[24] Because volume is a critical quality attribute and incorrect volume measurements would have a high impact on product quality, the decision consequence would be high. However, for this example, a manufacturer, as a part of release testing, would measure fill volume on a representative sample for each batch. Measuring fill volume through release testing would reduce the AI model influence, and therefore the model influence would be determined to be low. Given that the decision consequence is deemed high and the model influence is deemed low with the stated mitigations, the model risk for this COU is medium.

Assessing model risk is important because the credibility assessment activities used to establish the credibility of AI model outputs, which are described in step 4, should be commensurate with the AI model risk and tailored to the specific COU.

4. Step 4: Develop a Plan to Establish AI Model Credibility Within the Context of Use

Step 4 of the framework is to develop a plan to establish the credibility of AI model outputs. For the purposes of this guidance, such plans will be referred to as credibility assessment plans. Subsections 4.a and 4.b discuss general considerations and assessment activities related to establishing and evaluating the credibility of AI model outputs that can be included in such plans. These general considerations and assessment activities are not meant to be exhaustive, and some may not be applicable for all AI models and contexts of use.

Whether, when, and where the plan will be submitted to FDA depends on how the sponsor engages with the Agency, and on the AI model and COU.[25] For example, the plan could be described in a formal meeting package,[26] or another appropriate engagement option (see section

[24] For further information, see the guidance for industry *Allowable Excess Volume and Labeled Vial Fill Size in Injectable Drug and Biological Products* (June 2015).

[25] The Agency recognizes that certain uses of AI occur outside of contexts with established meeting options. Specifically, in the context of postmarketing pharmacovigilance, certain documentation (e.g., processes and procedures) is not generally submitted to the Agency but is maintained according to the sponsor's standard operating procedures and made available to the Agency upon request (e.g., during an inspection). In such cases, sponsors may choose to complete all the steps outlined in the guidance without seeking early engagement with the Agency. Sponsors remain responsible for compliance with statutory and regulatory requirements, including postmarketing safety surveillance and reporting requirements, regardless of the technology utilized.

[26] See the draft guidances for industry *Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products* (September 2023) and *Product-Specific Guidance Meetings Between FDA and ANDA Applicants Under GDUFA* (February 2023). When final, these guidances will represent FDA's current thinking on these topics. Also see the guidances for industry *Formal Meetings Between the FDA and ANDA Applicants for Complex Products*

IV.C below). The risk-based credibility assessment framework envisions interactive feedback from FDA concerning the assessment of the AI model risk (step 3) as well as the adequacy of the credibility assessment plan (step 4) based on the model risk and the COU. Accordingly, FDA strongly encourages sponsors and other interested parties to engage early with FDA to discuss the AI model risk, the appropriate credibility assessment activities for the proposed model based on model risk and the COU. Although detailed information on all the credibility assessment activities described in subsections 4.a and 4.b may not be available or necessary to include at the time of early engagement with FDA, the proposed credibility assessment plan about which the sponsor engages with the Agency should, at a minimum, include the information described in steps 1, 2, and 3 (i.e., question of interest, COU, and model risk) and the *proposed* credibility assessment activities the sponsor plans to undertake based on the results of those steps. In early discussions with the Agency, the proposed credibility assessment activities in the credibility assessment plan might be more high-level with a more detailed credibility assessment plan drafted after the iterative process.

As noted previously, the potential use of AI in the drug product life cycle is broad and rapidly evolving. Therefore, the activities that may be used to establish credibility of AI model outputs should generally be tailored to the specific COU and commensurate with model risk. For example, the performance acceptance criteria should be more stringent and described to FDA in more detail for high-risk models compared to low-risk models.

a. Describe the model and the model development process

The sponsor's credibility assessment plan submitted to FDA for early consultation should include the sponsor's *proposed* credibility assessment activities based on the question of interest, COU, and model risk. As noted previously, early descriptions of those activities may be high-level with further details provided after Agency feedback. In addition, for certain low-risk models, FDA may request minimal information in the categories described below. For high-risk models, FDA may request all of the information in the categories described below and additional information, as applicable, depending on the COU.

i. Describe the model

Sponsors and other interested parties should include the following information in the credibility assessment plan, as applicable, for each AI model used:

- An explanation of each model used including, but not limited to, descriptions of:
  - Model inputs and outputs
  - Model architecture (e.g., convolutional neural network)

under GDUFA (October 2022), and *Formal Meetings Between the FDA and Sponsors or Applicants of BsUFA Products* (August 2023). For information on combination product meetings, see the guidance for industry and FDA staff *Principles of Premarket Pathways for Combination Products* (January 2022).

313
314 – Model features[27]
315
316 – Feature selection process and any loss function(s) used for model design and
317 optimization, as appropriate
318
319 – Model parameters[28]
320
321 • A rationale for choosing the specific modeling approach
322
323 ii. Describe the data used to develop the model
324
325 For the purposes of this guidance, the data used to develop the model are generally composed of
326 training and tuning data[29] (collectively, development data) as part of the development stage.
327 Training data are those used in procedures and training algorithms to build an AI model,
328 including to define model weights, connections, and components. Tuning data are typically used
329 to evaluate a small number of trained AI models. More than one tuning dataset may be used as
330 part of the tuning process. The tuning process involves exploring various aspects for model
331 development, including different architectures or hyperparameters. The tuning phase happens
332 before the testing phase of the AI model and is part of the development stage (see subsection 4.b
333 for information on the testing phase).[30]
334
335 The performance of an AI model relies heavily on the datasets used to train and tune the model.
336 Therefore, the data used to develop the AI model should be fit for use, which means the data
337 should be both relevant (e.g., includes key data elements and sufficient number of representative
338 participants or sufficient data that is representative of the manufacturing process or operation)
339 and reliable (i.e., accurate, complete, and traceable).
340

[27] For the purposes of this guidance, a model feature is a measurable property of an object or event with respect to a set of characteristics. Features can include clinical measurements, demographics, and clinical imaging data. Features play a role in training and prediction. In the clinical development example discussed in section IV.A.1, model features include baseline demographic characteristics and lab values for trial participants (adapted from *ISO/IEC 23053:2022 - Framework for Artificial Intelligence Systems Using Machine Learning*).

[28] For the purposes of this guidance, a model parameter is an internal variable of a model that affects how it computes its outputs. Examples of parameters include the weights in a neural network and the transition probabilities in a Markov model (adapted from *ISO/IEC 22989:2022 Information Technology - Artificial Intelligence Concepts and Terminology*).

[29] Although some in the AI and ML communities sometimes use the term *validation* to refer to the tuning data and the tuning process, FDA does not use the word validation in this context.

[30] The definitions of *training and tuning data* for the purposes of this guidance are consistent with how those terms are discussed in the guidance for industry and FDA staff *Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions* and the draft guidance for industry and FDA staff *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations*.

Commensurate with model risk, sponsors and other interested parties should describe the data management practices for the development datasets (i.e., training and tuning datasets) and characterize the development datasets. These descriptions may help identify potential limitations of the data, including potential sources of algorithmic bias[31], and the appropriate credibility assessment activities to support use of the AI model for a particular COU. Sponsors and other interested parties should include the following information in the credibility assessment plan, as applicable:

- Describe (1) the development datasets, including how the development datasets were split into training, tuning, and any additional subsets and (2) the specification of which model development activities were performed using each dataset.

- Describe how the development data have been or will be collected, processed, annotated, stored, controlled, and used for training and tuning of the AI model. In addition:

  – Provide the rationale for choosing the specific development dataset(s).

  – Explain how labels or annotations were established.

- Describe how the development data is fit for the COU.

  – Explain how the development data is relevant (e.g., includes key data elements and sufficient number of representative participants or sufficient data that is representative of the manufacturing process or operation) and reliable (i.e., accurate, complete, and traceable).

- Describe whether development data are centralized (e.g., use of federated learning).

- Describe which model development activities were performed using each dataset.

iii. Describe model training

Commensurate with model risk, sponsors and other interested parties should include the following information on model training in the credibility assessment plan, as applicable:

[31] Data management is also an important means of identifying and mitigating bias and promoting health equity. Algorithmic bias is a potential tendency to produce incorrect results in a systematic, but sometimes unforeseeable, way due to limitations in the training data or erroneous assumptions in the machine learning process. For example, during training, models can be over-trained to recognize features that are unique to specific patient subpopulations, that have little to do with generalizable patient anatomy, physiology, or condition, which can result in AI bias in the resulting model. Additionally, for example, underrepresentation of certain populations in datasets could lead to overfitting (i.e., data fitting too closely to the potential biases of the training data) based on demographic characteristics, which can impact the AI model performance in the underrepresented population.

- Describe how the model was trained including, but not limited to, the:
  - Learning methodology (e.g., supervised, unsupervised).
  - Performance metrics used to evaluate the model, such as the area under the receiver operating characteristic (ROC) curve, recall or sensitivity, specificity, positive/negative predictive values (PPV/NPV), true/false positive and true/false negative counts (e.g., in a confusion matrix), positive/negative diagnostic likelihood ratios (PLR/NLR), precision, and/or F1 scores. All performance estimates should be provided with confidence intervals.
  - Techniques employed to prevent over- or under-fitting (e.g., regularization techniques).
  - Training hyperparameters (e.g., the loss function and learning rate).
- Specify whether a pre-trained model (or multiple pre-trained models) was used.
  - If a pre-trained model was used, specify the dataset that was used for pre-training and how the pre-trained model was developed and/or obtained.
- Describe the use of ensemble methods.
- Explain any calibration of the AI model (e.g., fine adjustment to the output of a trained model aimed at improving accuracy and/or repeatability).
- Describe the quality assurance and control procedures of computer software (including its toolboxes and packages) and how version changes were tracked.

b. Describe the model evaluation process

This subsection describes the evaluation of the fully trained model to assess the adequacy of the model performance for the intended COU on test data. Test data are those used to characterize the performance of the model. Test data should be independent of the development data and should not be shown to the algorithm during training. Instead, test data are used to assess the AI model's performance after training. Like development data, these data should be fit for use.

Commensurate with model risk, sponsors and other interested parties should include the following information in the credibility assessment plan regarding model evaluation, as applicable:

- Describe how the test data have been or will be collected, processed, annotated, stored, controlled, and used for evaluating the AI model.

- In addition:

  – Specify how data independence was achieved between development (training and tuning data) and test data. For example, data independence could have been achieved using data from a different clinical trial or health care system or data acquired using different batches or products.

  – If there was any overlapping use of data between the development stage and the testing phase, provide an explanation of how those data were used and a justification for why that use was appropriate.

  – As relevant, describe the reference method used to create the test data, and include a summary of the reference method's performance.

- Describe the applicability of the test data to the COU. This issue is important because, for example, when prediction models are developed using historical development data, the AI model may not perform as well in the COU if the development data are different from the data encountered in the deployed environment used in the COU. This phenomenon is sometimes referred to as *data drift*.

- Describe the agreement between the model prediction and the observed data, using test data that should be independent of the development data.

- Provide the rationale for the chosen model evaluation method(s) and explain the applicability of the evaluation methods to the modeling method used and to the COU. If the COU involves a "human in the loop," ensure that the evaluation methods consider the performance of the human-AI team, rather than just the performance of the model in isolation.

- Describe the performance metrics used to evaluate the model, such as the area under the receiver operating characteristic (ROC) curve, recall or sensitivity, specificity, positive/negative predictive values (PPV/NPV), true/false positive and true/false negative counts (e.g., in a confusion matrix), positive/negative diagnostic likelihood ratios (PLR/NLR), precision, and/or F1 scores, including the optimization methods used (e.g., use of a gradient descent). All performance estimates should be provided with confidence intervals. In addition:

  – Specify the process by which the uncertainty and confidence level of model predictions were estimated. If relevant, include any other descriptions or metrics that quantify confidence or uncertainty. Information regarding the uncertainty of model output is important because it helps interpret model outputs. Repeatability and/or reproducibility studies may help quantify the uncertainty associated with model outputs.

- Describe the limitations of the modeling approach, including potential biases.

- Describe the quality assurance and control procedures for code verification, including resolution of any errors or anomalies (e.g., user-generated codes are error-free, calculations are accurate).

5. Step 5: Execute the Plan

This step involves executing the credibility assessment plan. As discussed in step 4, discussing the plan with FDA prior to execution may help (1) set expectations regarding the appropriate credibility assessment activities for the proposed model based on model risk and COU and (2) identify potential challenges and how such challenges can be addressed.

6. Step 6: Document the Results of the Credibility Assessment Plan and Discuss Deviations From the Plan

Step 6 involves documenting the results of the credibility assessment plan and any deviations from the plan. This step generally occurs during the execution of the credibility assessment plan and should include a description of the results from steps 1 through 4.

The results of the credibility assessment plan should be included in a report. For the purposes of this guidance, this report is referred to as a credibility assessment report. The credibility assessment report is intended to provide information that establishes the credibility of the AI model for the COU and should describe any deviations from the credibility assessment plan as outlined in step 4. During early consultation with FDA (described in step 4), the sponsor should discuss with FDA whether, when, and where to submit the credibility assessment report to the Agency. The credibility assessment report may, as applicable, be (1) a self-contained document included as part of a regulatory submission or in a meeting package, depending on the engagement option, or (2) held and made available to FDA on request (e.g., during an inspection). Submission of the credibility assessment report should be discussed with FDA.

7. Step 7: Determine the Adequacy of the AI Model for the Context of Use

Based on the results documented in the credibility assessment report, a model may or may not be appropriate for the COU. If either the sponsor or FDA determine that model credibility is not sufficiently established for the model risk, several outcomes are possible: (1) the sponsor may downgrade the model influence by incorporating additional types of evidence in conjunction with the evidence from the AI model to answer the question of interest; (2) the sponsor may increase the rigor of the credibility assessment activities or augment the model's output by adding additional development data; (3) the sponsor may establish appropriate controls to mitigate risk; (4) the sponsor may change the modeling approach; or (5) the sponsor may consider the credibility of the AI model's output inadequate for the COU; therefore, the model's COU would be rejected or revised in an iterative fashion.

B. Special Consideration: Life Cycle Maintenance of the Credibility of AI Model Outputs in Certain Contexts of Use

For the purposes of this guidance, *life cycle maintenance* refers to the management of changes to AI models whether incidentally or deliberately, to ensure the model remains fit for use over the drug product life cycle for its COU. Life cycle maintenance of AI models is a set of planned activities to monitor and ensure the model's performance and its suitability throughout its life cycle for the COU.

As mentioned in section III, life cycle maintenance of the credibility of AI model outputs is important because a model's performance can change over time or across deployment environments. While the use of AI to support regulatory decision-making for drugs is typically assessed on locked data and information produced by an AI model at a given point in time, there are instances where the use of AI models extends over the drug product life cycle, and life cycle maintenance of the credibility of AI model outputs is critical. For example, life cycle maintenance of the credibility of AI model outputs is important for the application of AI modeling in the pharmaceutical manufacturing phase of the drug product life cycle.[32]

AI-based models may be highly sensitive to variations or changes in model inputs, for example, because they are data-driven and can be self-evolving (i.e., capable of autonomously adapting without any human intervention). Model performance metrics should be monitored on an ongoing basis to ensure that the model remains fit for use and appropriate changes are made to the model, as needed. The level of oversight for a model over its life cycle should be risk-based (i.e., commensurate with the model risk and the COU). Due to the evolving nature of AI models, sponsors should anticipate inherent, model-directed changes and the need to identify and evaluate those changes, as well as any intentional changes to the model over the drug product life cycle.

A risk-based approach[33] for life cycle maintenance may help sponsors assess the impact of a change or changes to the AI model performance. For example, in pharmaceutical manufacturing, it is important that changes to the AI model or changes in manufacturing that may impact the performance of the AI model be evaluated by the manufacturer's change management system within their pharmaceutical quality system (e.g., newly available manufacturing data or information, new signals requiring manual changes in the model, model-directed changes that may impact AI model performance).[34] The impact of a model change may be determined based

[32] Life cycle maintenance of AI modeling may be important during other phases of the drug product life cycle including, but not limited to, the application of AI modeling in the postmarketing phase. Section IV.B is focused on AI modeling in the pharmaceutical manufacturing phase as an example.

[33] See footnote 12, which provides additional references discussing FDA's application of benefit-risk principles when assessing the safety, effectiveness, and quality of a drug.

[34] See the ICH guidance for industry *Q10 Pharmaceutical Quality System* (April 2009). For further information, visit FDA's web page Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice

on factors such as model risk (see step 3 in section IV.A.3) and change in model performance. Depending on the extent of the change and its impact on model performance, some steps in the credibility assessment plan may need to be re-executed, including retraining and retesting the model for the COU. Additionally, depending on the impact of model change (i.e., if the model change impacts model performance), the change should be reported to the Agency in accordance with regulatory requirements.[35]

In general, detailed plans about life cycle maintenance (e.g., model performance metrics, the risk-based frequency for monitoring model performance, and triggers for model retesting) should be made available for review as a component of the manufacturing site's pharmaceutical quality system, with a summary included in the marketing application for any product or process-specific models, in accordance with regulatory requirements.[36] FDA recommends that the level of detail regarding life cycle maintenance of the AI model be commensurate with model risk.

Sponsors may also choose to use tools outlined in the ICH guidance for industry *Q12 Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle Management* (May 2021), such as established conditions and comparability protocols (referred to as postapproval change management plans), which leverages increased product and process knowledge. Sponsors may propose model-related elements to be considered established conditions, along with a plan to manage changes to these established conditions over the drug product life cycle. By including such plans in the marketing application, sponsors may prospectively obtain input from the Agency regarding management of such changes, including which changes would not require submission to the Agency prior to making modifications.

C. Early Engagement

As noted previously, FDA strongly encourages sponsors and other interested parties to engage early with FDA to (1) set expectations regarding the appropriate credibility assessment activities for the proposed model based on model risk and COU and (2) help identify potential challenges and how such challenges may be addressed.

Various options can be used to engage with the Agency, depending on how the sponsor or other interested parties intend to use the AI model. To discuss the use of AI in connection with a specific development program, sponsors may request an appropriate formal meeting (e.g., Initial

---

Regulations at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/quality-systems-approach-pharmaceutical-current-good-manufacturing-practice-regulations

[35] For example, as appropriate for the application type, such update would generally be made as a postapproval change in accordance with section 506A of the FD&C Act and 21 CFR 314.70 (for human drugs), 21 CFR 601.12 (for human biological products), or 21 CFR 514.8 (for animal drugs). The mechanism for postapproval notification of changes to models can be determined on the basis of the following two factors: (1) impact of the change on model's performance and (2) impact of the change on product quality.

[36] See 21 CFR 314.50 and 601.2.

Targeted Engagement for Regulatory Advice (INTERACT) on CBER/CDER Products, Pre-Investigational New Drug Application (Pre-IND)).[37]

Table 1 provides a list of various *other* engagement options depending on the intended use of the AI model. Where the meeting request covers a specific development program under an investigational new drug application (IND) or a pre-IND, sponsors should include the IND or pre-IND number and notify the relevant review team of the meeting request.

Table 1. Engagement Options Other Than Formal Meetings

Engagement Option	Intended Use of AI Model	Contact Information
Center for Clinical Trial Innovation (C3TI)	Sponsor is interested in discussing the use of AI in clinical trial designs with CDER before formally submitting them to their investigational new drug (IND) application	Email CDER C3TI program at CDERclinicaltrialinnovation@fda.hhs.gov
Complex Innovative Trial Design Meeting Program (CID)	Sponsor is interested in using AI in novel clinical trial designs	For details about how to apply for the CID program, please see http://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program FDA encourages sponsors to send an email to CID.Meetings@fda.hhs.gov to provide notification that your CID meeting request application has been submitted.
Drug Development Tools (DDTs) and Innovative Science and Technology Approaches for New Drugs (ISTAND)	Sponsor or other interested party is interested in qualifying a drug development tool that uses AI, such as use of AI-based algorithms to evaluate patients, adjudicate endpoints, or analyze clinical trial data	Email CDER Biomarker Qualification Program at CDER-BiomarkerQualificationProgram@fda.hhs.gov Email CDER Clinical Outcome Assessment Qualification Program at COADDTQualification@fda.hhs.gov Email CDER and CBER Animal Model Qualification Program at

[37] See footnote 26.

Engagement Option	Intended Use of AI Model	Contact Information
		CDERAnimalModelQualification@fda.hhs.gov Email CBER DDT Qualification Programs (includes Biologics Biomarkers and Clinical Outcome Assessments) at CBER-DDTQualificationProgram@fda.hhs.gov Email the ISTAND Pilot Program at ISTAND@fda.hhs.gov
Digital Health Technologies (DHTs) Program	Sponsor or other interested party is interested in using an AI-enabled DHT used in the context of a drug development program	To discuss general feasibility for a proposed DHT, or for those with general questions about the potential use of their DHT, email DHTsforDrugDevelopment@hhs.fda.gov
Emerging Drug Safety Technology Program (EDSTP)	Sponsor or other interested party is interested in using AI in pharmacovigilance (PV) EDSTP is specifically focused on the use of AI in PV for postmarketing activities; it is part of CDER's multifaceted approach to enhance mutual learning of where and how specific innovations, such as AI, can best be used throughout the drug product life cycle	EDSTP is not an avenue to seek regulatory advice on compliance with pharmacovigilance regulations. Questions about a specific development program should be addressed through other channels. Please contact AIMLforDrugDevelopment@fda.hhs.gov with the subject line "EDSTP" for more information.
CDER's Emerging Technology Program (ETP) and CBER's Advanced Technologies Team (CATT)	Sponsor or other interested party is interested in uses of AI in pharmaceutical manufacturing	Early engagement with the ETP or CATT is highly encouraged before submitting a regulatory application or implementing an AI technology for drug or biological product manufacturing. Requests and proposals may be sent by

Engagement Option	Intended Use of AI Model	Contact Information
		email: For CDER regulated drugs CDER-ETT@fda.hhs.gov, and for CBER regulated biological products Industry.Biologics@fda.hhs.gov, include "CATT" in the subject line.
Model-Informed Drug Development Paired Meeting Program (MIDD)	Sponsor is interested in using Model-informed drug development using AI	Sponsors with a pre-IND or an IND who are considering the application of MIDD approaches to the development and regulatory evaluation of medical products in development should email MIDD@fda.hhs.gov with "MIDD Program Meeting Package for CDER" (CDER applications) or "MIDD Program Meeting Package for CBER" (CBER applications) in the subject line.
Real-World Evidence (RWE) Program	Sponsor or other interested party is interested in using AI in a study using real-world data to produce RWE	For more information on the CDER, CBER, or OCE RWE programs, please visit each center's web page or contact CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov

589