May 1, 2023

Dockets Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: Docket No. FDA–2023-N-0487: Discussion Paper: Artificial Intelligence in Drug Manufacturing, Notice: Request for Information and Comments

Dear Sir/Madam:

The Biotechnology Innovation Organization (BIO) thanks the Food and Drug Administration (FDA or Agency) for the opportunity to submit comments regarding the discussion paper on **Artificial Intelligence in Drug Manufacturing.**[1]

BIO is the world's largest trade association representing biotechnology companies, academic institutions, state biotechnology centers and related organizations across the United States and in more than 30 other nations. BIO's members develop medical products and technologies to treat patients afflicted with serious diseases, to delay the onset of these diseases, or to prevent them in the first place.

I. General Comments

BIO greatly appreciates and supports the FDA's efforts in seeking stakeholder feedback to assist in better informing the Agency's evaluation to improve future guidance related to Artificial Intelligence (AI) in Drug Manufacturing. BIO provides several additional areas for FDA's review and further consideration:

- BIO recommends CDER continue to closely coordinate with the CDRH Digital CoE on AI and other related digital technology initiatives to better align the two centers.
- BIO considers the ICH Q8/Q9/Q10 Points to Consider model definitions reasonable but the definition of "medium impact models" is not clear. This definitional ambiguity could lead to confusion when attempting to apply to AI-based models which could present issues with identifying appropriate post-marketing filing classifications. BIO recommends development and publication of use cases that demonstrate how AI medium impact models (e.g., for process control) could be

[1] **Discussion Paper: Artificial Intelligence in Drug Manufacturing, Notice; Request for Information and Comments** https://www.fda.gov/media/165743/download

managed under the pharmaceutical quality system with minimal or no regulatory reporting of changes.

- BIO recommends that guidances for the use of AI in manufacturing, to the extent possible, should be focused on general principles. Ideally, there should not be different expectations for different applications of AI-based models in the manufacturing process if they are an equal impact level (i.e., two different sets of expectations for two medium impact models carrying out different functions).
- BIO strongly recommends the use of protocols (e.g., Post-Approval Change Management Protocols (PACMPs)) as a mechanism to submit, review, and maintain models. BIO believes that increased frequency in reporting model changes will likely lead to many deviations and ultimately restrict the use of these models.
- BIO encourages the Agency to clearly define common terminology descriptions and include case studies, such as: digital twin / digital shadow / digital model; supervised and un-supervised learning, Process Analytical Technology (PAT).
- BIO recommends that the Agency consider the creation of platform technology master files for third parties to maintain certain proprietary information or processes.

II. SPECIFIC COMMENTS: Responses to FDA Questions

1. What types of AI applications do you envision being used in pharmaceutical manufacturing?

BIO agrees that AI has multiple applications in pharmaceutical manufacturing, most of these AI applications are being used in control and monitoring of manufacturing operations. This includes advanced process-control strategies, continued process verification, and ultimately real-time process optimization and automated operation and management of manufacturing. AI and the recent growth of computational power, paired with development of robust models and sophisticated analytics, has the potential to handle the enormous quantity of data generated through the manufacturing process.

The FDA discussion paper covers monitoring, identification and controlling opportunities for drug manufacturing. An additional benefit of AI's application to drug manufacturing may be in its ability to make predictions.

AI predictive capabilities based on historical data could be leveraged for manufacturing. For example, an AI model could inform and provide insight into when to change a specification, when to expect a deviation, or when to increase control before any physical deviation occurs based on the historical datasets and previous deviations analysis. BIO recommends predictive capabilities be addressed in the updated version of the discussion paper as outlined below.

AI Applications in Direct Product Control Including Advanced Process Control

- Visual Inspection Systems for defect detection of products (vials, syringes) via the use of continuous learning throughout the operations process to integrate learning from analysis of production samples into an improved algorithm.
- Statistical Process Control (SPC) by AI model or predictive monitoring of process and product performance (open loop control based on key process end point predictions, like process models for Multivariate Statistical Process Control (MSPC) using AI).
- Adaptive process parameter control within unit operations.
- AI based modeling to support analytical procedures for in-process or final product testing.

AI Applications in CMC Development and Process Design/Scale-up

- Predictive stability, or the use of AI modeling for shelf-life prediction for accelerated product development or to assess process changes that require stability data.
- Process modeling (unit-operations based or overall process train) for the identification of Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs) to target experimentation and the development of more robust processes.

AI Application in Product Quality Monitoring, Process Monitoring, Trend Monitoring

- Process troubleshooting and root cause analysis – AI can be applied to existing manufacturing processes to model the input/output relationship and support process troubleshooting and root cause analysis for unplanned events.
- Enabling continuous process verification – e.g., Active Learning (AL) and Machine Learning (ML) based solutions come equipped with contextualization capabilities and deep learning to mitigate issues with data integration, standardization, and analysis which will enable a Continuous Process Verification (CPV) environment.
- Predictive models for quality attributes for raw materials, final products, or intermediates.
- AI models in analytical procedures, such as particle size analysis (e.g., image analysis software for particle size analyzer) that can use a form of AI (machine learning: Convolutional Neural Network) to capture real-time data on the size and shape of powders and bulk solids.
- Digital twins that use AI for more direct process interventions or control.

AI Applications in Monitoring (less or no direct product quality impact)

- Make use of “big manufacturing data” to improve manufacturing processes (e.g., transfer of data and knowledge from various sources help manufacturers improve the batch drug-making process through optimized resource utilization.)
- Models for preventive and predictive maintenance (e.g., Equipment).
- Predicting key process indicators (KPIs and product quality attributes (PQAs) to optimize the production process (e.g. automated process control systems analyzes the vast scale of data obtained during the production process and applies advanced data analytics and machine learning to derive critical insights and enable manufacturers to identify bottlenecks in the process, leading to more cost effective and optimized production).

- Use of digital twins that use AI to process simulations for optimization, evaluation of planned adjustments, and to introduce or qualify new equipment and systems to leverage existing knowledge and performance to right size validation requirements.

BIO anticipates AI will continue to be applied for uses outside the quality system including root cause analyses and preventative maintenance. Within the quality system, we expect AI to continue to be used to support batch release together with product testing or for parametric release (e.g., in association with PAT). Further, we could see applications for batch dynamic set-point changes for biologics (advanced process control).

In addition to the above recommendations, BIO has identified other potential applications of AI to include:

- Detection of abnormality in real time from vision systems (observing human tasks) and heuristics or similar observing signals from digitized processes.
- Automated scheduling of laboratories and manufacturing to avoid stock out scenarios and optimize replenishment.
- Computer System Validation (CSV) and other validation types including testing of boundary cases and accelerating validation delivery.
- The replacement of traditional dosage form methodologies with predictive models using *in vivo* data (e.g., to remove analog of acid dissolution, bioequivalence dissolution, Drug Metabolism and Pharmacokinetics (DMPK)).
- Prediction on cumulative impact of change based on big data principles of past change and/or ratification of real-life effect of change with low threshold signals.
- Detection of signals across similar datasets (e.g., correlations between lab investigation, deviation, complaints, vendor complaints, and analytical data indicating undiscovered impacts and trends).
- Projection of design space of New Product Introduction (NPI), especially related to dosage form characteristics from material properties, inputs, and other similar processing operations.
- Digital twins using ML/hybrid models with closed loop control for enabling autonomous platforms that are capable of self-optimization, defining design space and conducting critical assessments.
- Optimization of New Product Introduction (NPI) with a minimum number of manufacturing batches to tackle multiple objectives at once (e.g., yield, impurity reduction, cost, sustainability targets).
- AI for automated calibration and maintenance of models.

- Optical Character Recognition-Natural Language Processing (OCR-NLP) to extract information from PDF documents.
- Common Standardized Auto ML hybrid model pipeline for process-related work across business verticals Mid / Late Phase Development, Primary & Secondary), with fully integrated ontology and digital twin (where applicable).
- The development of new models to mimic manual actions and anticipate production results.
- PAT to enable real time release and real time process monitoring.
- Use of digital twins for manufacturing processes and analytical assays.
- Real-time release supported by in-line, on-line or at-line measurements.
- Use of algorithms to identify abnormal data patterns and focus human review on higher risks rather than data controlled through data validation routines.
- Autonomous monitoring of stability data, including OOT and trend detection.
- Advanced Process Control using PAT sensors as data input for fine.
- Process monitoring to detect drifts from typical batches and cycles or "golden envelope", allowing the real time setting of warning thresholds, to replace CPP control via alarms by control via data review.
- Automatic boundary tests for validation and assistance in the selection of testing level definitions in a Computer Software Assurance (CSA) approach.

BIO applauds industry for the wider use of AI/ML models across the drug development lifecycle. Broadly speaking, the intended use of AI/ML models is to reduce variability, increase focus on product quality and patient safety, and reduce delivery timelines of products to patients. BIO members intend to implement risk-based static or locked algorithm AI/ML models across business functions, as suggested by FDA. In addition, BIO recommends the Agency provide additional clarity and context via guidance and other interactive forums for when and how to implement self-adaptive learning models on GxP-regulated computerized systems to leverage the maximum potential of the technology.

2. **Are there additional aspects of the current regulatory framework (e.g., aspects not listed above) that may affect the implementation of AI in drug manufacturing and should be considered by FDA?**

The discussion paper for the regulatory framework for the usage of AI/ML model in Software as Medical Device (SaMD) appears to be more focused on medical devices and its fit for intended use. BIO recommends additional guidance be provided for the use of both static and adaptive learning models intended for drug manufacturing organizations.

BIO also notes that AI/ML models require large sets of data to train and test which require maintenance and retention of data, i.e., data retention guidelines for GxP records maintenance. BIO recommends that appropriate guidance be provided for data retention requirements to ensure that intelligence and learnings for the AI/ML are not lost with industry-applicable data retention guidelines.

BIO request further insight and details regarding any specific regulatory requirements for resource training and skills to implement and monitor AI/ML models for GxP use. BIO would also like to understand if there are regulatory expectations on when a model is measured to be fit for the intended use, level of acceptability, model tuning, and other model parameters.

Beyond the implementation examples/use cases listed above, the following aspects of the existing framework may affect implementation of AI:

<u>Existing GMP Framework</u>

- There are existing guidelines for models (e.g., near infrared spectroscopy (NIR) guidelines for partial least squares (PLS) and principal component analysis (PCA) models, draft guidelines for use of models in device development), which need to be aligned with guidelines for AI models and provide clarity on what elements of existing guidelines apply to AI and what is new to AI.
- A framework (AI specific or derived from existing guidance) for risk-based AI classification (e.g., similar to approach published in ISPE paper).
- Particular aspects for AI to address may include documentation in dossier and life cycle management - in dossier and/or only in PQS, especially for future self-learning AI models

<u>General Guidance</u>

- The publicly visible regulatory discussion on the use of AI and ML in drug development is often focused on applications for “big data” such as real-world evidence and clinical trial data modeling, where specific consideration is given to challenges like data ownership, privacy, and, ethics which are not as prevalent in the AI application in manufacturing. Considering this, a distinction of regulatory guidance for manufacturing and CGMP versus such applications would be valuable.
- One common element in the discussion on the use of AI in “Big Data” is the “Explainability and Transparency”, expectation of documentation on understanding and interpreting how an AI model is working. This is frequently difficult due to the specific nature of algorithms (neural networks, deep learning etc.) and may stifle

the AI model deployment. Hence, BIO recommends other elements of ensuring model quality and performance be used, similar to the regulatory discussion on multivariate modeling in spectroscopic methods.

Other Aspects of Regulatory Framework

- Specifics on acceptable practices on data management, including specifics on the technologies industry can rely on, without compromising and in fact demonstrating data integrity principles.
- Increased focus on global initiatives beyond US (e.g., ICH or similar mechanisms).
- Discussion on quality requirements such as robustness, controllability, explicability, maintainability, conformance, predictability, reliability, and accuracy.
- AI/ML standard pipeline for common manufacturing/process related capabilities.
- Further discussion on how a dataset has been modified/impacted by an AI since this aspect of the technology could become more challenging over time.
    - Specify which uses will have regulatory considerations and which will not (i.e., use of cloud application should not have regulatory impact).
    - Validation and maintenance strategy for AI models/predictive models.
    - Opportunities to use protocols in place of discrete submissions to submit and maintain high-impact AI-based models in filings (e.g., PACMPs or related vehicle to describe the process by which a company maintains a model and keeps it in a validated state for process control in place of individual filings at every change.
    - For pooled data sets, it is currently unclear how FDA will determine applicability or how industry should ensure that datasets are maintained according to FDA standards. BIO recommends the Agency further explore the possibility of a DMF-like mechanism in submitting such data by a third party (e.g., a platform technology master file).

3. **Would guidance in the area of AI in drug manufacturing be beneficial? If so, what aspects of AI technology should be considered?**

BIO agrees that there is growing need for guidance for early adopters of AI/ML applied to pharmaceutical manufacturing. BIO supports future guidance for industry that would appropriately frame and answer AI-related questions in manufacturing processes. Such guidance should consider multiple aspects of AI technologies, such as:

- Terminology Alignment and Clarification: There is great opportunity for terminology alignment and clarification. Differences in definitions throughout the industry have caused substantial confusion. From a

regulatory perspective, it will be beneficial for both the Agency and industry if the Agency worked with industry to distinguish regulatory language from descriptions of scientific or engineering principles and practices.

Standards for Developing and Validating AI Models: BIO agrees there are limited industry standards and FDA guidance available for the development and validation of models that impact product quality. This lack of industry standards can create challenges in establishing the credibility of a model for a specific use

Data generation and managements for AI models: Applicants may need clarity regarding regulatory compliance for AI-generated data (e.g., which data needs to be stored and/or reviewed and how loss of these data would impact future quality decisions such as product recalls). Further, applicants may need additional clarity for data sampling rates, data compression, or other data management approaches to ensure that an accurate record of the drug manufacturing process is maintained."

- Validation and Verification of Continuously Updated AI Control Models: Applicants may need clarity on:

  (a) The expectations for verification of model lifecycle strategy and the approach for FDA's examination of continuously updated AI control models during a site inspection.

  (b) The expectations for establishing product comparability after changes to manufacturing conditions introduced by the AI model, especially for biological products.

- Regulatory Flexibility and International Harmonization
  - Technical guidelines will be important to guide development efforts; however, it is critical that these guidelines offer flexibility (especially given the highly dynamic nature of this technology) and are globally harmonized to the greatest extent possible. Reliance on pre-existing global standards (e.g., ISO) is also recommended, where possible.
  - Guidance on expectations for dossier content would be welcome. For AI models a highly burdensome reporting framework will dramatically reduce the practicality of introducing and maintaining these models. Therefore, it is strongly recommended that even for high impact models, the Agency rely on initial assessments of model credibility and justifications for the

intended use while relying on submitted protocols to maintain AI models in the PQS. Guidance on the development, submission, and maintenance of these protocols for AI based models would be beneficial and could rely on many elements of ICH Q12.

- Risk-based Approaches for Deployment of AI Models:
  - There is a need for risk based AI classification for GMP. For example, there are ongoing efforts to use AI for predictive maintenance of equipment, or scheduling. Further insight and clarity as to the extent AI, directly or indirectly, will be subject to CGMP requirements, compared to for example modeling applications directly linked to product control strategies or visual inspection systems.
  - Review/revise current Computer System Validation/Assurance (CSV) or Equipment Qualification (EQ) guidelines by incorporating AI or modeling aspects. It will be helpful to see guidelines on how model validation is executed under the current CSV and EQ structure.
  - Additional guidance is also needed on risk-based approach to data storage requirements and cybersecurity expectations for cloud-based solutions, especially considering that many AI models or data sets will be hosted in the cloud.

- Guidance on Steps That Lead to AI Model Deployment: As stated in the discussion paper, the existing guidelines on process models even without the use of AI is limited, so in general, the field of process models may benefit from further clarification to consider aspects specific to AI type models, including:
  - Data Collection – Label, Ingest, Aggregate: Expectations on the use of open-source libraries while developing custom ML/DL algorithms.
  - Data Preparation – Clean, Partition, Scale, Augment
  - Feature Engineering – Selection, Transformation, Creation (Encoding), Extraction
  - Model Development – Including guidelines on how to treat any exceptions that may not be covered in the training data set but may occur as a deviation in manufacturing.
  - Model Validation or Standard for Model Performance Assessment – Guidance clarifying whether it is acceptable to recycle validation data sets for related model building and to what extent, and whether it is acceptable to use historical data for model validation.

- Model Documentation - Clarity on what constitutes an AI model-the theoretical model, the code and model parametrization, the data used to train the model.
- Documentation for the Regulatory Dossier – For AI specifically, the documentation of the "inner workings of the model" (i.e., connecting inputs to outputs). Key discussion topics on AI include, "Explainability" and "Transparency" and the amount of information to include on these topics in the dossier.
- Lifecycle Management – in dossier and/or in PQS, especially related to "learning" algorithms.

- Guidance on "Learning" or Self-updating" Algorithms; Guardrails: Can the model be continuously improved/retrained or should this happen iteratively, so that each model parametrization can be referenced and linked to a passed product that was controlled with that model?

- Consideration of Applications of Natural Language Processing

- Consideration of Future Computing Platforms i.e., Cloud Computing and Quantum Computing. Use of AI for rethinking traditional (i.e., 3 batch validation) concepts of process performance qualification.

- Considerations for advanced process control applications (APCs)

- Considerations related to visual systems for error detection

- Considerations related to the GMP vs non-GMP uses of AI, and related criteria.

Additionally, guidance should further discuss what specifically constitutes an unexplainable "black box" model as opposed to a potentially acceptable explainable "white box" model. Are there specific algorithm limitations / recommendations, e.g.:

- Should a model algorithm be changed for performance reasons, what are the acceptable options: neural nets vs. random forest vs. SVM vs. gradient boost vs. linear regression, PCA, PCR, hybrid (if so, what allowable blends, if applicable), any validation R squared, RMSE bounds / limitations / considerations?
- Must SHAP rankings be present and available to explain the model?
- Must results include both +ve and –ve variances?
- Can eigenvectors and n-dimensionality be used?
- Under what scenarios would the algorithms (and combinations thereof) be acceptable?

4. **What are the necessary elements for a manufacturer to implement AI-based models in a CGMP environment?**

One overarching element is the global understanding of how AI works and where it can be incorporated. Not all problems can be solved with AI and AI is not meant to be implemented in every facet of a manufacturing process. This educational component regarding AI capabilities will be important to have to successful implement within the manufacturing process.

AI models' efficacy depends on the model itself and on the datasets entered into the model. To implement AI-based models in line with CGMP requirements it is important that manufacturers utilize labelled, structured, and normalized datasets. These data requirements can be regrouped under what's commonly called "data management and standardization principles". These principles should also include a part on data safety, security, and integrity.

BIO recommends the Agency consider the following elements:

- With increased connectivity across upstream and downstream processes across business functions and the drug development lifecycle, it may become increasingly important to identify and document system(s) of records that feed AI-based model(s).
- Guidance based on the intended use and risk-based model maintenance would be required. Applying a rigorous and robust maintenance model not considering the intended use can negate any benefits of using these models.
- Guidance around feature engineering, data quality, and data standardization. The guidance should specifically consider how data standardization will play a role in static model versus adaptive learning models.
- Guidance around the use of proprietary and open-source AI libraries.
- AI Governance strategy with clearly defined AI Process and Risk Management.
- How to apply a risk management approach to each phase of ML model lifecycle including data strategy in ML model development (ML model qualification starts at data collection phase via cleaning, identifying patterns and making predictions); augment the training data set periodically, monitor model performance and upgrade as necessary.

    - ML validation framework with model classification involving design control and autonomy levels – risk including the level of control desired, from requiring subject matter expert's approval to fully automation.
    - Strong documentation on internal PQS documents and procedures, as required in the regulatory dossier (risk-based guidance needed on this) and related life cycle management.
    - Cross organization upskilling training on AI Process Management, Validation and ML model risk assessment understanding.
    - For Active Process Control models, clarification as to what are the expected relationship to quality management systems. For example, would a change in the process operating set point trigger a change control process in manufacturing? If so, how can this be sustainable when many such changes are made based on evolving datasets?
    - Understanding the multi-dimensional data (e.g. (mechanistic) model data, vs IoT plant/equipment parametric data, vs PAT data, vs traditional QC methods) and the ability to combine and rank on-the-fly to determine appropriate source of truth and cause for discrepancies.
    - Modular/risk-based approach to scale validation (e.g., not GMP/Non-GMP), but more granular based on potential for patient/product impact and/or compliance impact. Thereby focusing validation resource on areas of import and acceleration of model refinement.
- Understanding the data to continuously and automatically validate the projection of the model against the measured reality in order to refine the model.
- Clear understanding of how an ICH Q10 framework for the PQS is applicable to AI models and what constitutes an acceptable period review of such a model.
- Defining the impact of AI on the product (and/or GxP decision) and being able to explain why the impact is under control through an appropriate risk-based approach (e.g., if product can be impacted but its quality is verified by other means later, impact on patient is reduced at a minimum while still improving the manufacturing process).
- Regulatory framework around AI/ML model should be established with guidance on reportable changes (AR) and prior-approval ones (CBE-30, PAS) vs. those that can be carried out in the PQS.
- Expansion of the use of QbD concepts (Design Space) as licensed established conditions, allowing real-time adjustment of manufacturing processes.
- Expectations related to site audits will be critical and especially important given recent Mutual Recognition Agreements (MRAs).
- Understanding of expectations for responsibilities over maintenance of third-party datasets, quality agreement elements etc.

BIO also highlights that a full branch of the QMS would need to be defined and dedicated to support of AI. Points to consider include:

- How and when would e-signatures be incorporated into AI output?
- How is the 4-eye principle applicable? (Is AI considered the first set of "eyes"?)
- Under what conditions could GMP documentation be generated by AI (SOPs, COAs, QP declarations) with reduced human oversight?

5. **What are common practices for validating and maintaining self-learning AI models and what steps need to be considered to establish best practices?**

The application of adaptive learning models in the pharmaceutical manufacturing industry is still nascent but should not limit regulatory agencies from providing relevant guidance on the use of such technologies.

BIO agrees there are limited industry standards and FDA guidance available for the development and validation of models that impact product quality. This lack of industry standards can create challenges in establishing the credibility of a model for a specific use.

Multiple validation methods for AI models are reported in the literature. These methods include:

- **Simulation**: In simulations validation methods, the system is validated in a clearly artificial environment (virtual or real) mimicking the actual deployment environment.
- **Trial:** Trial validation methods in a real environment aim to validate the system by using the system as it would be used, in the final deployment environment.
- **Model-centered**: Model-centered validation method is a set of validation methods that focus on ensuring the correct functioning of the used AI model. This type of validation is often data-centered and can be seen as having trust in the system by ensuring that the model works.
- **Expert Opinion**: Expertise required to conduct the validation is not present in the team responsible for the system but is obtained from the outside.

There is a growing need around industry standards for validating self-learning AI models in drug manufacturing. BIO supports future activities in this space that would define risk-based validation methods commonly supported by the industry.

An additional point to establishing best practices in the choice of validation models is that the re-evaluation of the AI model after any major updates could require validating the model

again. This re-evaluation will need to be included in future documents describing industry standards.

General basic rules for development, validation, and lifecycle management of models will apply, i.e., data selection, separate data sets for calibration/training and validation/verification, robustness considerations., updates, lifecycle management, etc. A starting point for AI models could be the existing guidelines for spectroscopic models.

In developing self-learning modules BIO notes that learning rules must be clear and balanced with Root Mean Square Error of Prediction (RMSEP), that the model is robust against outliers, and that model validity is constantly monitored. BIO recommends the Agency assess how a sponsor should assess a model is fit for the intended purpose, the reliability of the data, the validated of the self-learning system remains withing the set acceptance criteria to ensure absence of conflict. As noted, to support constant monitoring, regular checks should be implemented couple with continuous mathematical verification. All results of the AI should be explainable and understandable.

In terms of regulatory documentation considerations BIO notes that lifecycle management within PQS or retroactively (annual report) is necessary as cycle time of self-learning and improvement would be defined by submission and approval processing times and therefore essentially rendered unworkable and defeating the technical advantage and purpose of self-learning AI models. For APC, a framework with expected stress tests to be conducted to ensure models robustness and validation would be valuable. For other applications, there may be sufficient guidance based on published examples if regulators accept these approaches.

6. **What are the necessary mechanisms for managing the data used to generate AI models in pharmaceutical manufacturing?**

BIO recommends guidance to be published on mechanisms used to procure and use data for model training and testing. Guidance around the use of open-source AI libraries and propriety 3rd party closed AI models must be established.

Many already known mechanisms for data usage in a GMP environment apply to AI: data are rationalized, stored and version controlled for AI models. Missing data, outliers and how we treat them need to be established. For the noted topic areas for AI models (large volume, cloud-based applications, and IoT), it will be helpful to clarify the applicability and interpretation of existing guidelines vs. creating additional guidelines.

Other points more specific to data used for AI applications include the following:

- Ensure appropriate training data labeling to avoid bias.
- Tools to assess criticality of data (critical data elements vs non-critical data elements) based on potential impact may be needed. Impact can include but is not limited to the need for establishing model comparability to future versions, enough sampling rate to reconstruct same AI decisions, ensure that data is not corrupted, and demonstrate that data hasn't been manipulated (e.g. encryption for example for cloud solutions).
- Data standardization, considering that data for AI/ML use cases can be utilized from several different sources (data warehouses, lakes, cloud storage, databases).
- Data can have different levels of uniqueness and related criticality - e.g. visual inspection algorithms based on product specific data but also on "generic" image data from databases that help with filtering noise, reduce dimensionality and fine-tune the model. Use and documentation of "generic" data is important.

In addition, it is foreseeable that AI applications or models will be provided by 3rd party vendors. It is key to understand the expectations to ensure that 3rd parties follow applicable regulations (e.g., for cloud services) as these are often sold as black-box approaches.

The FDA should consider the following in describing how data integrity principles be demonstrated without compromising the flexibility and versatility of the most sophisticated approaches:

- Criticality of data governance to ensure source data quality.
- Single source of truth where necessary for inter-table relationships.
- Application of data models to become source independent.
- Long-term development of shared data standards across industry.
- Knowledge management to bring together disparate sources with non-obvious foreign key relationships (e.g., not dependent on absolute relationship management).
- Digitalization of knowledge and documentation sets (e.g., the end of documents and a shift to datasets for methods, specifications, etc).

BIO recommends the Agency consider that this will depend upon the model impact (e.g., high impact AI models would need to be managed under GMPs, low impact models could be handled in a manner like CMC data developed to support product development).

- Data strategy and governance: common language is required across industry, CMOs, and regulators.

In terms of the data itself, BIO recommends that the data needs to be:

- Generated (sensors, equipment generating precise estimates of critical parameters);
- Extracted (equipment connection, Networked Devices and Machines);
- Integrated (Data storage technology: data warehouse/Data Lake);
- Structured (augmented by Data governance, respecting ALCOA+);
- Treated to get analyzed, meaning and insight (statistics, modelled, cloud computing, algorithmic, AI, ML…); and
- Quickly accessible (insight interface).

In order to manage and analyze the data a common language is required across industry, CMOs, and regulators. The data, itself, must also be representative of the routine activities and cover most cases. In addition, the data used to train the model must be different than the data used to validate it, while clear threshold values must be defined and require manual action when reached.

Further, the stringency of the data management should depend on the model. A clear understanding of the data models, coming from the original raw data (which may be proprietary), extracted data for model use (which should be in open formant and contain reduced information/meta-data as compared to the raw) and the impact of this reduced set for the model in needed. There is also a need for a clear understanding of the data flow and the use of common ontologies. Extreme care must be taken for the data to be structured in an identical manner (as far as possible) within an organization, so that the data scientists can easily retrieve/classify/purge the datasets and that different sets of different data can be gathered/compared/aggregated, etc. Process business mappings are a good practice to help choose the relevant data and to ensure the accuracy, traceability, and reliability of the AI models used.

7. **Are there other aspects of implementing models (including AI-based models) for pharmaceutical manufacturing where further guidance would be helpful?**

BIO recommends guidance be provided on the use of COTS AI/ML models. Specifically, guidance is needed on the regulatory expectations for supplier assessment and quality agreements for AI/ML models used in manufacturing operations where computer systems are integrated into manufacturing equipment and analytical instruments.

Continuously learning AI systems that adapt to real-time data may challenge regulatory assessment and oversight. It may be challenging for manufacturers to assess when to submit a regulatory notification of changes to an AI model.

BIO agrees with FDA's statement that Applicants may need clarity on: (a) the expectations for verification of model lifecycle strategy and the approach for FDA's examination of continuously updated AI control models during a site inspection, and (b) expectations for establishing product comparability after changes to manufacturing conditions introduced by the AI model, especially for biological products.

CDRH's approach for AI incorporated medical devices is efficient to pair with the iterative functions of AI and the continuous learning capabilities. Such an approach could be applied to drug manufacturing where an applicant could describe a package of potential changes and implement the changes when needed based on the collection of sufficient data without resubmitting each time a package to FDA. Additionally, FDA needs to advance innovative mechanisms for evaluating technology outside product approvals. Any substantial acceleration in the pace of implementation of innovative technology requires FDA to engage earlier and more broadly in considering the suitability of novel enabling technologies.

As stated in the discussion paper, the existing guidance on process models - even without the use of AI - is limited, so in general the field of process models could benefit from further clarification (currently it appears that for these types of process models the FDA is applying draft guidance for use of models in devices from Dec 2021,Docket Number: FDA-2021-D-0980). The implementation of process modeling could be enhanced by guidance on the development, validation, and documentation aspects of AI. Global harmonization, especially with EMA in this field is desirable. As AI models might be developed often by 3rd parties, some sort of "validation label" that allows use of those models without further scrutiny by the user will be very helpful.

Further guidance would be helpful on the use of AI/ML for real-time release. Specifically, a review of the absolute requirements to achieve full real time release through parametric or PAT based methodologies and the validation approach versus either traditional method equivalence or (preferred) to DMPK/in vivo characteristics throughout development lifecycle at FTIH into commercial production. Instigation of parametric release is likely to operate according to CQAs linked to true patient effect. Common methodologies today try to prove equivalency to an analogue (traditional) method; and due to non-comparative methodologies, both methods and specifications are not suitable for equivalence.

If the AI is used only to build the model, it means that the acceptance criteria can be defined and measured. This then allows the use of traditional methods to validate the system. Hence, guidance on the development and use of "soft-sensors" - CQA not chemically / physically measured but inferred from other values / measurements.

8. **Are there aspects of the application of AI in pharmaceutical manufacturing not covered in this document that FDA should consider?**

BIO recommends the guidance speak to the importance of establishing ethical standards on the intended use of AI/ML models and the ability to use structured versus unstructured data models from both internal data resources and open-source AI libraries.

BIO recommends that the Agency discuss to clarify how it will use AI/ML models for thw purpose of regulatory inspections and audits of pharmaceutical manufacturers. For example, consider the following:

- ChatGPT and use of Large Scale Language Models in a GMP environment are not mentioned but will likely be seen in applications (documentation and other uses)
- Aspects of Ethics in the use of AI for GMP (or rather, delineation and potentially non-applicability of "ethics concerns" raised in other types of big data, like patient, clinical trial etc.)
- Use in stability/shelf-life modeling
- The need for more powerful computing platforms for the large data sets needed for AI e.g., Quantum Computing might be necessary as well and considerations on how current and future computing platform regulations apply.

The Agency should provide more added clarity and context about how the landscape has evolved across the industry and the relationship to current practices that are either regarded obsolete or still standing would help. A description of the expectation for the roles involved (e.g., quality assurance) might be helpful.

Cloud computing is supportive of AI but not only used in that context and the type of risks are quite different than the usage of AI. Guidance focusing on production and areas bound to the Quality Control laboratory and associated automation would also be valuable. A framework around the laboratory of the future is needed and would complement this initiative.

It would be helpful for the Agency to consider additional discussion of error-free workflow, whereby monitoring systems can guide operators, analysts and other GMP critical task without pre-training (i.e., training is conducted through guided workflow and error detection with no opportunity for digression or error). This requires AI and vision systems to fully detect and critically evaluate abnormal actions to ensure no breach of GMP standards for manufacture or testing and enable the critical parameters to be met. This will

have a major impact on labor capability, availability, and opportunity to remove the cost of poor quality (COPQ), therefore avoiding drug shortages.

BIO recommends that the Agency consider linking to relevant ICH guidelines and to discuss consideration for ICH M4Q CTD dossiers and how the information is linked to the model impact.

Sincerely,

/s/

Camelia Thompson, Ph.D.

Senior Director, Science and Regulatory Policy

Biotechnology Innovation Organization