

**PDA Global Headquarters**  
Bethesda Towers,  
Suite 600  
4350 East West Highway  
Bethesda, MD 20814 USA  
TEL: +1 (301) 656-5900  
FAX: +1 (301) 986-0296

**PDA Europe gGmbH**  
Am Borsigturm 60  
13507 Berlin  
Germany

**OFFICERS**

*Chair*

**Susan Schniepp**

*Chair-Elect*

**Anil Sawant, PhD**

*Secretary*

**Emma Ramnarine, PhD**

*Treasurer*

**Melissa Seymour, MBA**

*Immediate Past Chair*

**Jette Christensen, PhD**

*President & CEO*

**Glenn E. Wright**

**DIRECTORS**

**Bettine Boltres, PhD**

**Cristiana Campa, PhD**

**Javier Camposano, MBA**

**Cylia Chen Ooi**

**Mirko Gabriele, PhD**

**Marc Glogovsky**

**Andrew Hopkins**

**Stephan O. Krause, PhD**

**Ivy Louis, MBA**

**Amy McDaniel, PhD**

**Mathias Romacker**

**Osamu Shirokizawa**

09 August 2023

Dockets Management  
Food and Drug Administration  
5630 Fishers Lane, Rm 1061  
Rockville, MD 20852

Re: Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products (Docket Number: FDA-2023-N-0743)

Dear Sir or Madam:

The Parenteral Drug Association (PDA) is submitting the attached responses on behalf of a group of our members regarding FDA's discussion paper on Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products. The paper provided the industry with FDA perspectives on this issue and PDA's members have collaborated to collectively reply to the questions posed by the agency. In the attached comments, the team offers insight and suggestions that may assist the agency in formalizing and finalizing guidance related to Artificial Intelligence and Machine Learning.

The group of members wish to emphasize that regardless of the guidance in place around the use of these new technologies, there remains a need for human oversight of the automated processes, the algorithms employed, and the results generated.

PDA is a non-profit international professional association of more than 10,000 individual member scientists having an interest in the fields of pharmaceutical, biological, and device manufacturing and quality. These comments have been prepared by a group of volunteers with expertise in pharmaceutical and biopharmaceutical manufacturing with the aim of aligning on best practices and policies to ensure patient safety and continuity of drug supply.

If you have any questions, please do not hesitate to contact me via email at [eaton@pda.org](mailto:eaton@pda.org).

Sincerely,



Josh Eaton  
Senior Director, Scientific and Regulatory Affairs

cc: Glenn Wright, President and CEO

## Responses to “Discussion of Considerations and Practices for AI/ML in Drug Development”

### (1) Human-led governance, accountability, and transparency

*a. In what specific use cases or applications of AI/ML in drug development are there the greatest need for additional regulatory clarity?*

- i. The use of big data in drug development is well suited to AI/ML. However, governance around the metadata must ensure that data integrity is maintained to support its use in future regulated product development.
- ii. Data pre-processing actions (manual), identification of relevant variables, algorithm development, algorithm selection and fine-tuning, AI model monitoring and real-time AI model execution are areas where specific guidance is required.
- iii. There is a great need in the context of automated visual inspection and the use of deep neural networks, which are quite different from the other AI/ML models used in continuous manufacturing and for which explainability is very difficult. Specific guidance may be required depending on the particular algorithm(s) in use (e.g., the rationale for the model, expected results for the use of that model).
- iv. Based on the following examples, this team sees the need for guidance on developing an overall AI/ML development, management, and control plan:
  1. Validation of the AI/ML experience to ensure equal representation of all test subjects during clinical trials to ensure drug product works as intended using good clinical study management practices. This is particularly important when an unplanned drug reaction occurs at the human level to defend selection criteria and demographics.
  2. Unit-level traceability (the ability to serialize down to the container level using RFID), represents a key digital enabler for a range of Pharma 4.0™ initiatives and can power an AI/ML model. Considered an IOT, there is no clarity on regulatory agencies' expectations nor how data from these tools would be managed through cloud platforms and services.
  3. The automatization of the full process (i.e., from receipt to shipment and interfaces in ERP and LIMS to MES that support batch release). This would include dataset construction, rules for selection for training/validation/testing based on quantity and diversity (metrics), operation (based only on engineering/control run or also on real product) and requirements for operators (is it training or validation?), qualification of equipment (is this based only on engineering/control run or also on real product?).

*b. What does transparency mean in the use of AI/ML in drug development (for example, transparency could be considered as the degree to which appropriate information about the AI/ML model—including its use, development, performance, and, when available, logic—is clearly communicated to regulators and/or other stakeholders)?*

We believe the following methods of transparency for AI/ML are primarily intended for regulatory agencies to allow for adequate oversight and evaluation of AI/ML in drug development.

- i. During drug development, AI transparency would allow the logical decision tree the AI/ML model uses to come to a conclusion or generate data, to become available during audits and to scientists. We would need to ensure the model remains open,

visible, and traceable from one change to the next and able to determine over which generated dataset the decisions were made. This impacts common drug development phases as well as targeted cell and gene therapy drug development.

- ii. A clear Digital Strategy / Digital Management System should be required of each company, just as having a Quality Management System is required. This should spell out the AI/ML models, lifecycle, metadata, data types used in each model, etc.
- iii. Transparency means reliable documentation to standardize manual and automated processes: data clean-up and relevant variable identification, strategies to avoid overfitting, nulls, missing values, etc. Also, the applied criteria for algorithm selection, maintenance of the library versions used in the algorithm, fine-tuning of the algorithm, monitoring of the data drifting, AI model monitoring, etc. Additional guidance on how to accomplish this would be beneficial.
- iv. Transparency is the ability to explain and verify the model behavior based on the dataset provided for training. For Deep Neural Network the more effective way to verify and communicate the DNN behavior is using the heatmaps that show the contribution of each pixel in the image to the final decision. The main issue is how to demonstrate the completeness of the solution beyond the dataset, it needs statistical methods and appropriate metrics to verify it.

*c. In your experience, what are the main barriers and facilitators of transparency with AI/ML used during the drug development process (and in what context)?*

- i. Data Validation - The willingness or ability to articulate the rationale of using AI/ML. This requires a deep knowledge of the tools at hand and why they were selected in accordance with validation principles as they pertain to AI (e.g., as outlined in section 2.4.2 in the EMA "Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle"). An additional consideration is how to validate the AI/ML experience to ensure equal representation of all test subject during clinical trials to ensure drug product works as intended using good clinical study management practices. Especially when an unplanned drug reaction occurs at the human level defending selection criteria and demographics is necessary and would also need to consider trial diversity (for example, trial rollout from one country vs another).
- ii. Data Management - Data clean-up, transformation, data preparation are not always transparent or traceable. Algorithm selection, fine-tuning and model creation are tasks that can be easily reproduceable and assessed. The ability to export and encapsulate AI models in standard formats (like ONNX) are enablers and facilitators to ensure that AI Models do not work as black boxes.
- iii. Data Collection - In the context of Automated Visual Inspection (AVI), the main barriers are the data collection and labelling that require huge investment in man-hours and the lack of an effective metric to verify that the data collection is exhaustive enough for the model to train. Overfitting is always a big risk having models with such a large number of parameters and defective images that are so scarce that could not cover the full range of variability. The use of a Generative Adversarial Network (GAN) in association with human-in the loop labelling could be a way to alleviate the issue, provides guidelines and automatically identifies mistakes in manual labelling. Such a baseline would then be fine-tuned by transfer learning for each specific case following the AI development success stories happened in academic and other industrial sectors.

*d. What are some of the good practices utilized by stakeholders for providing risk-based, meaningful human involvement when AI/ML is being utilized in drug development?*

- i. The automatization of the full process: Scripts for data cleaning and data preparation, algorithm fine-tuning, data set for testing and training, model creation, real time outputs, etc. Everything should be automated in scripts in order to avoid repetitive actions based on human interaction.
- ii. In the AVI context, the human inspectors supervise the labelling of images for the training/testing Dataset and manually revise the rejects from the machine to verify the correctness of the result by comparing those images with the images stored by the machine. This is then verified by sampling (physical and images) the accepted images to look for new defects not recognized by the DNN that need to be included in the dataset for further training. An effective way for finding outliers is to run a second DNN/GAN as an anomaly detector trained on conforming products to learn the accepted quality that could trigger any anomaly for further investigation. The suggested deviation will be validated or rejected by the operator, like in the medical field, reducing the human effort and the risk of incorrect classification or missing detection.

*e. What processes are in place to enhance and enable traceability and auditability?*

- i. Data Integrity principles like ALCOA+, address the required traceability and auditability of the process related to data management. Nevertheless, similar guidelines are needed to maintain the traceability and auditability of the algorithm and model management process (e.g., can AI/ML system audit trails differentiate between actual data input vs interpretive data (filling in gaps) autogenerated by the system).

*f. How are pre-specification activities managed, and changes captured and monitored, to ensure the safe and effective use of AI/ML in drug development?*

- i. New data used to retrain models or to create new models could require resetting of the algorithms and even new algorithms and hence, new models. Model monitoring and model updates must follow a change control policy, driven by robust risk analysis.

## **(2) Quality, reliability, and representativeness of data**

*a. What additional data considerations exist for AI/ML in the drug development process?*

- i. Data drifting is a reality that happens in drug development. The reality is that when AI is deployed and it interacts with equipment, material, systems, and humans, the interaction itself predicates a change in the behavior of the full system. Monitoring the AI model output establishing the right acceptance conditions is a good practice, but it must be complemented with monitoring of data drift, which would be the root cause of potential deviation in the AI Model outputs.
- ii. Data leakage is a great concern in terms of contamination of real data with training/APS/engineering data. It is important for manufacturers to have a process in place to account for this potential occurrence.
- iii. Data augmentation is a popular practice to solve data scarcity and imbalance but needs to be controlled and documented. Synthesized data must be recognizable from original data to verify the AI does not introduce bias due to the characteristics of synthesized data.

- 
- iv. Setting limits tighter than those a human is capable of attaining calls into question all previous human inspections. Therefore, even if the AI/ML method can be set to its capability, the setting needs to be determined by human performance trending.

*b. What practices are developers, manufacturers, and other stakeholders currently utilizing to help assure the integrity of AI/ML or to address issues, such as bias, missing data, and other data quality considerations, for the use of AI/ML in drug development?*

- i. In the realm of CGT, we use “qualified” patient material/test subjects in our clinical studies. AI/ML can help us in the selection of qualified patients as it can screen patients for the specific cell types needed for the study as well as randomly select qualified patients for needed demographics to ensure all genders, ages, weights, heights, those with other predisposed factors are equally represented in the study and ensure successfully clinical trial results and safety interpretation occurs. AI/ML can also be used to determine if a test subject (qualified patient) is an outlier in the data map based on the study criteria and be used as justified rationale for not including that test subject’s data in the study.
- ii. The main practices currently implemented by AI developers are totally aligned with the existing good practices for software (where AI = software + data + calculations). Additionally, the GAMP5 v.2 based on risk assessment is a very good approach to include in the AI lifecycle since AI is a good tool to objectivize the risk assessment.
- iii. In the AVI context, data quality is assured by the experience of technical experts that know the source of variability, the nature of the defects, and the previous QC knowledge of the statistical relevance of deviations.
- iv. Transfer learning is seen as a promising way to reduce the need for large datasets for scarce products by transferring the knowledge acquired with similar products (freezing the input layers of the DNN model) and fine-tuning only the last layers of the model with real word defects (RWD) to fine-tune the classification. Transfer learning needs rules to establish the knowledgebase (bracketing approach to be defined) and metrics to document and validate it.

*c. What are some of the key practices utilized by stakeholders to help ensure data privacy and security?*

- i. AI/ML can help with random selection of test subject (qualified patient) lot numbers to eliminate patient privacy concerns currently experienced by CGT companies.
- ii. HIPAA and ISO 27001 enclose good practices and guidances to cover both aspects.
- iii. In the AVI context, most stakeholders have decided to use AI on premises without migrating the data out of their facilities. This is possible for large organizations, but for small enterprises a cloud-based infrastructure would be a better choice, however data security and privacy are considerations. To address this, data privacy and security is ensured by excluding the labelling metadata (anonymization) from training/validation/test datasets to avoid the risk of indirectly holding privacy information in the model and as a source of bias.

*d. What are some of the key practices utilized by stakeholders to help address issues of reproducibility and replicability?*

- i. The main good practice is to transform manual actions (like data cleaning or algorithm fine-tuning) into scripts which are well-documented and maintained.
- ii. In the AVI context, there are stakeholders relying on storing all the data and labels for reproducibility and replicability and DNN models for backup and recovery. This approach is demanding large investment in data storage capabilities, but currently there is no better approach to provide the same replicability.
- iii. Stochastic modelling effects like normalization require a larger dataset and robust performance metrics as related to small output variations between successive runs. Choosing metrics non-sensitive to class imbalance, like Matthew's correlation coefficient (MCC), is another important consideration to keep in mind in the project design.

*e. What processes are developers using for bias identification and management?*

- i. The concern with bias in AI algorithms is that AI/ML may not be able to fully eliminate bias if a certain population of test subjects all meet the case study/clinical trial criteria and fall into the same demographic criteria. Equally, there should be consideration to potentially include more representation from smaller demographic criteria if that population has an identified impact of the clinical trial. The aim is for AI to ensure equal representation across all demographics, include specified training datasets as programmed, or "red flag" the company if it is incapable of meeting certain demographic criteria.
- ii. A dedicated committee to oversee data management is essential to ensure that no biases are introduced in the data. Furthermore, there are statistical techniques to measure and calculate bias in data or samples not well represented.
- iii. Up to now it is based on human interpretation of the heat maps produced during validation and maintenance of the AI system. Tight integration of data and metadata is paramount for effective data management. ALCOA+ requires adoption of JSON or XML format, however AI/ML can interpret formats that are not human-readable and therefore a rationale for the usage of and a strategy for interpretation of AI/ML formats is required.

### **(3) Model development, performance, monitoring, and validation**

*a. What are some examples of current tools, processes, approaches, and best practices being used by stakeholders for:*

*i. Documenting the development and performance of AI/ML models that can be applied in the context of drug development (e.g., CONSORT-AI (Liu et al., 2020) and SPIRIT-AI (Cruz Rivera et al., 2020))?*

1. Qualifying AI algorithms for the expected intended use (as described in Artificial Intelligence Algorithm Qualification: A Quality by Design Approach to Apply Artificial Intelligence in Pharma; Manzano, Toni, et al. *PDA J Pharm Sci Tech* Jan 2021, 75 (1) 100-118; DOI: 10.5731/pdajpst.2019.011338)

2. The process should be challenged periodically to ensure that the model has maintained its validated state. The most used tools in the context of AVI are open-source tools and the embedded metrics available like precision, accuracy, sensitivity, specificity, F1, MCC scores, ROC, and AUC.
- 

*ii. Selecting model types and algorithms for a given context of use?*

1. Use of a decision tree to determine the conditions for the right algorithm selection (e.g. [choosing the right estimator](#): Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011)
  2. In the context of AVI, the most used AI model are the convolutional neural network (CNN) going from the most compact VGG to ImageNet depending on the target hardware, dataset dimension and execution cycle required.
- 

*iii. Determining when to use specific approaches for validating models and measuring performance in a given context of use (e.g., selecting relevant success criteria and performance measures)?*

1. Considerations associated with clinical and R&D applications and then the subsequent transfer to manufacturing processes. Depending on the stage-gate processes, validation may need to be aligned with clinical decisions for patient assessments and when this affects the manufacturing processes.
  2. The relevant success criteria must be defined immediately after describing the intended use. AI Models must be validated and properly monitored when they are involved in critical tasks.
- 

*iv. Evaluating transparency and explainability and increasing model transparency?*

1. Using standard formats for AI model encapsulation (like ONNX) is a good way to evaluate the transparency and explainability of the AI Models.
  2. Many deep learning explanation techniques are post hoc, which means they try to make sense of a trained neural network by examining its output and its parameter values. For instance, one technique to determine what a neural network sees in an image is to mask different parts of an input image and observe how these changes affect the output of the deep learning mode. This technique creates saliency maps that highlight the relevant features in the input images.
- 

*v. Addressing issues of accuracy and explainability (e.g., scenarios where models may provide increased accuracy, while having limitations in explainability)?*

1. AI model monitoring is a very good strategy to ensure that models are providing the expected accuracy. Nevertheless, the best way to ensure the accuracy is measuring the data drifting of the information fed into the model. Also, this strategy is more sensitive to potential deviations.
2. Out-of-distribution generalization is not a well-addressed concern, therefore the algorithms used must be checked and monitored consistently. The checks should include identifying anomalous data (i.e., data from domains that have not been seen during training) which falls outside of the validated state.

*vi. Selecting open-source AI software for AI/ML model development? What are considerations when using open-source AI software?*

1. Primary considerations are availability, regular updates, and good documentation in terms of initial installation and subsequent management of the software following any updates.
  2. Security of the code to prevent intentional adulteration of the data or algorithms.
- 

*vii. The use of RWD performance in monitoring AI/ML?*

1. RWD are used to monitor the production environment. Any deviation from the expected values triggers an alarm action that could require the retraining of the AI/ML model. Correlating data related to deviations (e.g., images, patient data) is an invaluable tool to improve knowledgebase.

*b. What practices and documentation are being used to inform and record data source selection and inclusion or exclusion criteria?*

- i. Metadata should be considered as part of a company's Digital Management System, with rationale applied to each dataset's metadata decision criteria.

*c. In what context of use are stakeholders addressing explainability, and how have you balanced considerations of performance and explainability?*

- i. An assessment provided by a dedicated data management committee is crucial to ensure a balanced consideration of performance and explainability.

*d. What approaches are being used to document the assessment of uncertainty in model predictions, and how is uncertainty being communicated? What methods and standards should be developed to help support the assessment of uncertainty?*

- i. Risk assessment tools are good mechanisms to support and manage uncertainty in models once the confusion matrix is continuously evaluated as soon as new data is integrated and outputs produced.