Incorporating AI into the Process Validation Lifecycle: A Two-Way Street

Alyssa Burke, Senior Validation Engineer

BW Design Group

Introduction

Artificial intelligence (AI) continues to evolve and propel businesses forward; however, the Life Science industry's adoption of these technologies is delayed in comparison with other industries. AI technologies have the ability to greatly reduce cost and time to market of products, especially when applied to and integrated with the process validation lifecycle. Due to the nature of the highly regulated industry, these technologies also introduce a number of challenges to the industry, particularly when it comes to validating these solutions. This presentation aims to demonstrate how developing AI solutions integrates with the process validation lifecycle, while demystifying some of the nuances around various AI solutions and terminology.

What's the Difference?

AI: artificial intelligence; broader concept of machines simulating human intelligence[2]

VS.

ML: machine learning; subset of AI where systems learn patterns from data and make decisions without explicit programming[2]

VS.

GenAI: generative AI; type of AI that specializes in creating new, original content[2]; i.e. ChatGPT

Algorithm: set of rules or procedures for solving a problem or task; ex: PID[2]

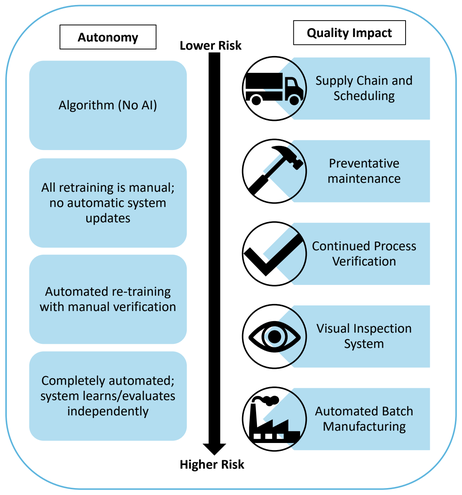
VS.

Model: result of applying an algorithm to training data; performs predictions by utilizing learned patterns

AI Solution Complexity Levels

Consider:

1. ***How autonomous is the system?***[1]
2. **What is the likelihood of quality impact on the final product/substance?**[1]



Incorporating AI into the Process Validation Lifecycle

The graphic below aims to demonstrate how the development and validation of an AI system aligns with the process validation lifecycle. In this example, the AI system is used for process optimization and continued process verification.

Process Development

Goal

Identify intended use of system, data sources, algorithms, and integration with drug process.

Actions

- Define system requirements
- Select AI model and data sources
- Identify CPPs and CQAs to be monitored
- Develop initial models and train them
- Perform initial exploratory analyses to evaluate accuracy and explore trends
- Specify acceptance criteria for model performance
- Perform risk analysis

Documentation

- User Requirement Specification (URS)
- Functional Specification (FS) and Design Specification (DS)
- Data Integrity Plan
- Risk Assessment
- Validation Master Plan

System Development

- Data collection from PAT and lab include temperature, pressure, mix speed, humidity, API concentration, potency, hardness
- Utilize AI model such as autoencoder to detect outliers, utilize explainable AI such as SHAP and LIME to determine most relevant factors
- Include data from development batches

Process Qualification

Goal

Demonstrate AI system performs reliably under expected manufacturing conditions.

Actions

- Verify correct installation of system
- Confirm compliance with data integrity and cybersecurity requirements
- Test system in controlled environment
- Validate algorithm performance against predefined acceptance criteria
- Conduct stress testing
- Deploy AI system in production

Documentation

- IQ/OQ/PQ
- SOP Development
- User acceptance testing
- Model Validation Reports
- Change control records
- Training records

System Validation

- Challenge model with new data set during process qualification
- Manually check if detected outliers are true outliers
- Evaluate 21 CFR Part 11 compliance
- Challenge target metrics for reconstruction error of autoencoder (low for normal points, high for outliers)
- Refine precision, recall, F-1 score limits for model

Continued Process Verification

Goal

Ensure system remains accurate, reliable, and compliant.

Actions

- Monitor model performance with key performance metrics
- Compare model outputs to actual outcomes
- Periodic retraining of the model
- Perform re-validation when model is updated
- Investigate any failures of the model
- Update risk assessments
- Conduct internal audits of the system
- Maintain records and change controls

Documentation

- Continuous monitoring reports
- APQR
- Retraining and revalidation protocols/reports
- Risk assessment
- Change control records
- Incident investigations and CAPAs

System Deployment

- In production: model alerts operator of any anomalies detected in real-time and indicates possible causes
- Retrain model every 6 months with updated data sets
- Investigate any false anomaly alerts and any missed anomalies
- Model automatically generates reports of its precision/recall and provides notification when they fall out of limits

Process Optimization: Simulated Case Study

A simulated case study was executed using randomly generated datasets by Python. Variable process parameters include Temperature, pH, Mix Speed, and API Quantity. Target attributes are Purity and Assay.

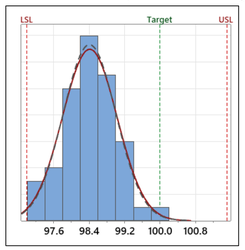


Figure 1: Process capability analysis for assay (%). Ppk = 0.77, Cpk = 0.79

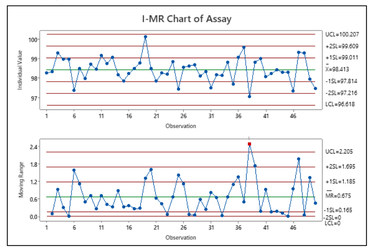


Figure 2: Control chart for assay (%) with one outlier on the Moving Range chart.

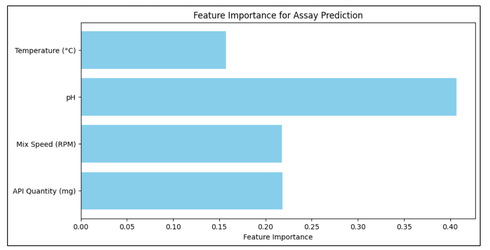


Figure 3: Feature importance plot for assay from Random Forest Model indicating pH has greatest impact on pH.

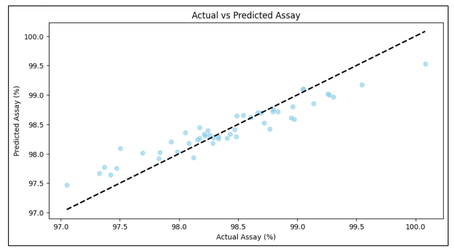


Figure 4: Plot of actual assay values compared to assay values predicted by the Random Forest model on a test data set.

Conclusions drawn from capability analysis and control charts:

- Normal distribution suggests stable process (Figure 1)
- Distribution is not centered around the target value indicating room for improvement (Figure 1)
- The low Ppk and Cpk values are a result of multiple data points very close to the lower specification limit.
- One outlier in the moving range chart indicates a large jump between two batches that may require additional investigation (Figure 2)

Next Steps:

- Design of experiments (DOE) with multiple development batches to determine what factors are impacting assay and adjust process as necessary to achieve results closer to target.

A random forest machine learning model was trained on an initial data set of 50 values and then evaluated on a different test data set to see if it can predict assay and purity from the four (4) parameters: temperature, pH, mix speed, and API quantity. Execution occurred in Python (Google Colab).

Results of model:

- Assay: Mean squared error = 0.06, R^2 score = 0.85
- Purity: Mean squared error = 0.00, R^2 score = 0.85
- Figure 4 demonstrates the accuracy of the model for assay. Predictive capabilities are shown to weaken at the extremes of the data pool indicating that additional data points at the high and low end may help improve performance of the model.
- Figure 3 indicates that pH has the greatest impact on assay (%)

Next Steps:

- Focus on optimizing pH levels to meet the target assay value, rather than varying all 4 variables and executing a full DOE.

Comparing Solutions for Stage 3 Validation (CPV)

How do advanced machine learning models compare to traditional CPV methods and multivariate statistics? Which option is the right solution for your process?

Autoencoder + SHAP/LIME[3]

- Type: AI/Machine Learning
- Strengths:
  - Detects non-linear trends and complex relationships in high dimensional data
  - Minimal human intervention
  - SHAP/LIME adds interpretability
- Cons:
  - Computationally intensive
  - Requires larger data set for training and validation
  - Requires additional validation due to black box AI components
- Best uses: Complex, non-linear anomaly detection

Multivariate Principal Component Analysis (PCA)

- Type: Multivariate statistics
- Strengths:
  - Well-established and easy to validate
  - Effective in reducing dimensionality
  - Helps identify correlated variables
- Cons:
  - Assumes linear relationships
  - Cannot detect non-linear patterns
  - Difficult to interpret principal components without linking to process variables
- Best uses: Identifying sources of process variability

Process Capability Analysis + Control Charting

- Type: Classic Statistics
- Strengths:
  - Simple and well understood
  - Requires minimal computational resources
  - Effectively detects large shifts in data
- Cons:
  - Cannot detect small or multivariate shifts
  - Lacks predictive capability
  - Does not adapt well to complex data
- Best uses: Simple process monitoring and quantifying capability

Conclusion

AI enhances process understanding, anomaly detection, and continued process verification; increasing efficiency, reducing costs, and improving product quality. To ensure compliance and reliability, organizations must establish robust validation strategies, maintain data integrity, and address regulatory expectations. By leveraging explainable AI and aligning model validation with traditional principles, pharmaceutical manufacturers can confidently implement AI-driven solutions while maintaining compliance. It is important to recognize that advanced tech solutions are not always the answer. Moving forward, professionals must develop a risk-based validation framework balancing innovation and compliance.

References and Acknowledgements

1. Erdmann, N., Blumenthal, R., Baumann, I., & Kaufmann, M. (2022, March). *AI maturity model for gxp application: A foundation for AI validation*. ISPE. https://ispe.org/pharmaceutical-engineering/march-april-2022/ai-maturity-model-gxp-application-foundation-ai
2. *Libguides: Artificial Intelligence: Openai, CHATGPT, llms, and more: Glossary of ai terms*. National University Library. (2025, February 25). https://resources.nu.edu/AI-at-NU/glossary
3. Antwarg, L., Miller, R. M., Shapira, B., & Rokach, L. (2019, March). *Explaining anomalies detected by autoencoders using ...* Arxiv. https://arxiv.org/pdf/1903.02407

This research utilized ChatGPT (OpenAI, 2025) through Microsoft Azure for language refinement and idea generation.

Statistical analysis was performed in Minitab Statistical Software 22.

Machine learning model execution and data generation was performed in Google Colab, utilizing Python libraries.

Let's Connect!
Alyssa Burke
Senior Validation Engineer
BW Design Group
Email: alyssa.burke@bwdesigngroup.com
Phone: +1 (845) 825-3431