

# AI MATURITY MODEL FOR GXP APPLICATION:

## A Foundation for AI Validation

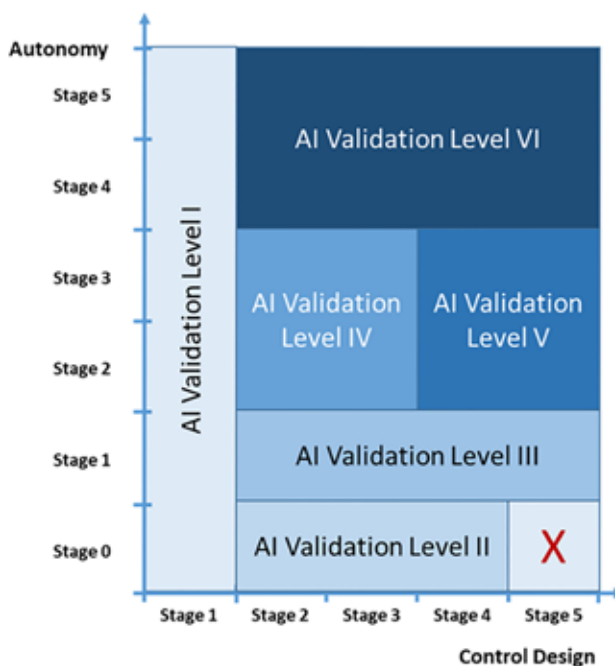
By Nico Erdmann, Rolf Blumenthal, Ingo Baumann, and Markus Kaufmann

Artificial intelligence (AI) has become one of the supporting pillars for digitalization in many areas of the business world. The pharmaceutical industry and its GxP-regulated areas also want to use AI in a beneficial way. Several pharmaceutical companies are currently running digital pilots, but only a small fraction follows a systematic approach for the digitalization of their operations [1] and validation. However, the assurance of integrity and quality of outputs via computerized system validation is essential for applications in GxP environments. If validation is not considered from the beginning, there is considerable risk for AI-based digital pilots to get stuck in the pilot phase and not move on to operations.

There is no specific regulatory guidance for the validation of AI applications that defines how to handle the specific characteristics of AI. The first milestone was the description of the importance and implications of data and data integrity on the software development life cycle and the process outcomes [2]. No life-science-specific classification is available for AI. There are currently only local, preliminary, general AI classifications that were recently published [3].

This lack of a validation concept can be seen as the greatest hurdle for successfully continuing digital products after the pilot phase. Nevertheless, AI validation concepts are being discussed by regulatory bodies, and first attempts at defining regulatory guidance have been undertaken. For example, in 2019 the US Food and Drug Administration published a draft guidance paper on the use of AI as part of software as a medical device [4], which demonstrates that the regulatory bodies have a positive attitude toward the application of AI in the regulated industries.

Figure 1: Maturity model.



### INTRODUCING A MATURITY MODEL

As part of our general effort to develop industry-specific guidance for the validation of applications that consider the characteristics of AI, the ISPE D/A/CH (Germany, Austria, and Switzerland) Affiliate Working Group on AI Validation recently defined an industry-specific AI maturity model. In general, we see the maturity model as the first step and the basis for developing further risk assessment and quality assurance activities. By AI system maturity, we mean the extent to which an AI system can take control and evolve based on its own mechanisms, subject to the constraints imposed on the system in the form of user or regulatory requirements.

**Table 1: Control design stages.**

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
The system is used in parallel to the normal GxP processes	The system is executing a GxP process automatically but must be actively approved by the operator	The system is executing the process automatically but can be revised by the operator	The system is running automatically and controls itself	The system is running automatically and corrects itself

**Table 2: Autonomy stages.**

Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Fixed algorithms are used (No machine learning)	The system is used in a locked state. Updates are performed by manual retraining with new training data sets	Updates are performed after indication by the system with a manual retraining	Updates are performed by automated retraining with a manual verification step	The system is fully automated and learns independently with a quantifiable optimization goal	The system is fully automated and self-determines its task competency and strategy

Our maturity model is based on the control design, which is the capability of the system to take over controls that safeguard product quality and patient safety. It is also based on the autonomy of the system, which describes the feasibility of automatically performing updates and thereby facilitating improvements.

We think that the control design and the autonomy of an AI application cover critical dimensions in judging the application’s ability to run in a GxP environment. We thus define maturity here in a two-dimensional matrix (see Figure 1) spanned by control design and autonomy, and propose that the defined AI maturity can be used to identify the extent of validation activities.

This article was developed as part of a larger initiative regarding AI validation. The maturity model is the first step. In fact, many other topics such as data management or risk assessment have to be considered in the validation of AI. The basic maturity model will have an influence on the risk assessment of the AI application.

In this article, we describe in detail which validation activities are necessary for AI systems with different control mechanisms and the varying degrees of autonomy that need to be investigated via critical thinking. The goal was to find clusters with similar validation needs across the entire area (see Figure 1) defined via the autonomy and control design dimensions.

## CONTROL DESIGN

Table 1 shows the five stages of the control design.

In Stage 1, the applications run in parallel to GxP processes and have no direct influence on decisions that can impact data integrity, product quality, or patient safety. This includes applications that run in the product-critical environment with actual data. The application may display recommendations to the operators. GxP-relevant information can be collected, and pilots for proof of concept are developed in this stage.

In Stage 2, an application runs the process automatically but must be actively approved by the operator. If the application calculates more than one result, the operator should be able to select one

of them. In terms of a 4-eye principle (i.e., independent suggestion for action on the one hand and check on the other hand), the system takes over one pair of eyes. It creates GxP-critical outputs that have to be accepted by a human operator. An example for a Stage 2 application would be a natural language generation application creating a report that has to be approved by an operator.

In Stage 3, the system runs the process automatically but can be interrupted and revised by the operator. In this stage, the operator should be able to influence the system output during operation, such as deciding to override an output provided by the AI application. A practical example would be to manually interrupt a process that was started automatically by an AI application.

In Stage 4, the system runs automatically and controls itself. Technically, this can be realized by a confidence area, where a system can automatically control whether the input and output parameters are within the historical data range. If the input data are clearly outside a defined range, the system stops operation and requests input from the human operator. If the output data are of low confidence, retraining with new data should be requested.

In Stage 5, the system runs automatically and corrects itself, so it not only controls the outputs but also initiates changes in the weighting of variables or by acquiring new data to generate outputs with a defined value of certainty.

To our knowledge, there are currently no systems in pharmaceutical production at level 4 or 5. Nevertheless, with more industry experience, we expect applications to evolve for applications at levels 4 and 5.

## AUTONOMY

Autonomy is represented in six stages (shown in Table 2).

In stage 0, there are AI applications with complex algorithms that are not based on machine learning (ML). These applications have fixed algorithms and do not rely on training data. In terms of validation, these applications can be handled similar to conventional applications.

**Table 3: Validation levels.**

Level	Description	Minimum Validation Activities and Requirements
I	Parallel (AI) CS <sup>a</sup>	No validation required
II	Classical non-AI CS	Validation of computerized system, but no dedicated focus on AI
III	Piecewise locked state <sup>b</sup> AI CS	In addition to the above requirements: <ul style="list-style-type: none"> <li>• Documented justification on why a model was selected</li> <li>• Training data verification</li> <li>• Model quality assurance after training</li> <li>• Input data monitoring in operation</li> <li>• Retraining procedures defined</li> </ul>
IV	Self-triggered learning AI CS with human operation and update control at all times	In addition to the above requirements: <ul style="list-style-type: none"> <li>• Monitoring of model quality in operation</li> <li>• Controlling quality KPIs<sup>c</sup> and notification process</li> <li>• Validation of the human factors (depending on control design) with regards to overrides, qualifications, and AI system acceptance</li> </ul>
V	Self-triggered learning AI CS with update control, but overall or sampled operation control only	In addition to the above requirements: <ul style="list-style-type: none"> <li>• Periodic re-test with defined test data set</li> <li>• Assurance of self-control</li> <li>• Control of AI system outcomes by samples for a defined, risk-oriented fraction, and adequate stratification of input/output instances</li> </ul>
VI	AI CS with autonomous learning	Validation concept currently under development

<sup>a</sup>CS means computerized system.

<sup>b</sup>Piecewise means that the system may be regularly or irregularly manually updated to another version but provides one exact output to an instance of input data within such a version.

<sup>c</sup>KPI means key performance indicators.

In stage 1, the ML system is used in a so-called locked state. Updates are performed by manual retraining with new training data sets. As the system does not process any metadata of the produced results by which it could learn, the same data input always leads to the generation of the same output. This is currently by far the most common stage. The retraining of the model follows subjective assessment or is performed at a regular interval.

In stage 2, the system is still operating in a locked state, but updates are performed after indication by the system with a manual retraining. In this stage, the system is collecting metadata of the generated outputs or inputs and indicates to the system owner that a retraining is required or should be considered, e.g., in response to a certain shift in the distribution of input data.

In stage 3, the update cycles are partially or fully automated, leading to a semi-autonomous system. This can include the selection and weighting of training data. The only human input is the manual verification of the individual training data points or the approval of the training data sets.

In stage 4 and stage 5, the system is completely autonomous with reinforced ML independently based on the input data.

In stage 4, the system is fully automated and learns independently with a quantifiable optimization goal and clearly measurable metric. The goal can be defined by optimizing one variable or a set of variables. In production, the variables could be the optimization of the yield and selectivity of certain reactions.

In stage 5, the system learns independently without a clear metric, exclusively based on the input data, and can self-assess its task competency and strategy and express both in a human-understandable form. Examples could be a translation application that learns based on the feedback and correction of its user. If the user suddenly starts to correct the inputs in another language, in the long term, the system will provide translations to the new language.

## VALIDATION LEVELS

The maturity levels can be clustered into six AI validation levels (see Table 3) and placed into the area defined by the dimensions of autonomy and control design (see Figure 2). The AI validation levels describe the minimum control measures necessary to achieve regulatory compliance of the systems at a high level. Detailed quality assurance requirements should be defined individually based on the categorization, given the intended use and the risk portfolio of the AI system.

Systems in AI validation level I have no influence on product quality and patient safety (and data integrity); therefore, validation is not mandatory. Nevertheless, for applications in this category, the human factor should not be underestimated. If a system is designed to provide advice and is running in parallel to the normal process for a prolonged amount of time, safeguards should be in place to ensure that the operator is handling results based on critical thinking and does not use these results to justify decisions.

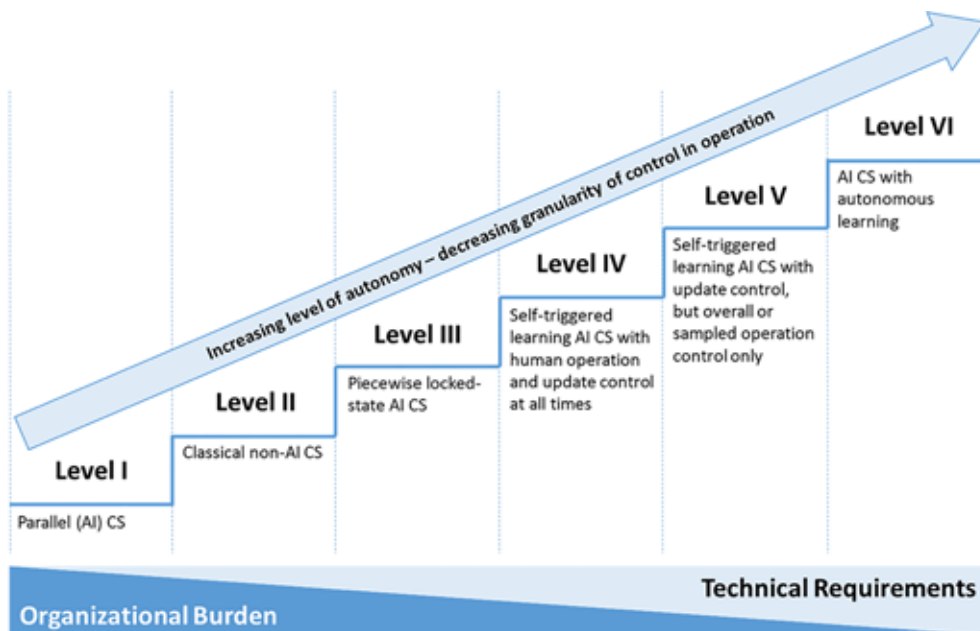
Systems in AI validation level II are AI applications that are not based on ML and therefore do not require training. The results are purely code-based and deterministic and can therefore be validated using a conventional computerized system validation approach.

Systems in AI validation level III are based on mechanisms such as ML or deep learning. They rely on training with data for the generation of their outputs. Systems in this category are operating in a locked state until a retraining is performed.

For the validation, AI-specific measures have to be performed that relate to the data model and the used data, in addition to the conventional computerized system validation. The integrity of the training data has to be verified. It needs to be verified that the data used for the development are adequate for generating a certain output and are not biased or corrupted. AI validation documents should cover the following aspects:

- A risk analysis for all extract, transform, load (ETL) process steps for the data
- Assessments of the data transformation regarding the potential impact on data integrity
- The procedures on how labels have been produced and quality assured

Figure 2: Representation of the validation levels with focus on business decision-making.



In addition, the model quality has to be verified during the development and operational phases. During development, it must be verified that:

- The selected algorithm is suitable for the use case
- The trained model technically provides the anticipated results based on the input data

In the operational phase, these additional aspects have to be considered and defined:

- Appropriate quality measures to monitor the model performance
- Required conditions to initiate retraining depending on model performance

For retraining, it is desirable that the input structures for the model input remain the same. Otherwise, a new assessment of the methodological setup of the development phase may be required.

To ensure that the system is only operating in a validated range, input data during operations have to be monitored. Furthermore, for systems in this category and above, transparency issues come into play, as the rationale for the generation of outputs based on different input data may not be obvious. For this reason, all available information should be visible to provide insight into the path to the outcome, and explainability studies (which aim to build trust in AI applications by describing the AI-powered decision-making process, the AI model itself, and its expected impact and potential biases) should be conducted to validate the decision-making process and provide explanations and rationales to any interested party.

Systems in AI validation level IV already inherit greater autonomy as varying aspects of the update process are automated, which can include the selection of new training data. For this reason, there is a strong need to focus on controlling key performance indicators that reflect model quality during operation. Model quality outputs should be monitored to ensure they are in the validated range. In addition, the notification process for cases, where the system requires a retraining or is operating outside of the validated range, must be confirmed.

Systems in AI validation level V have a greater process control. Therefore, stronger system controls have to be in place during the operation. This can be achieved by periodically retesting with defined test data sets. Furthermore, the self-controlling mechanism should be verified during the validation phase.

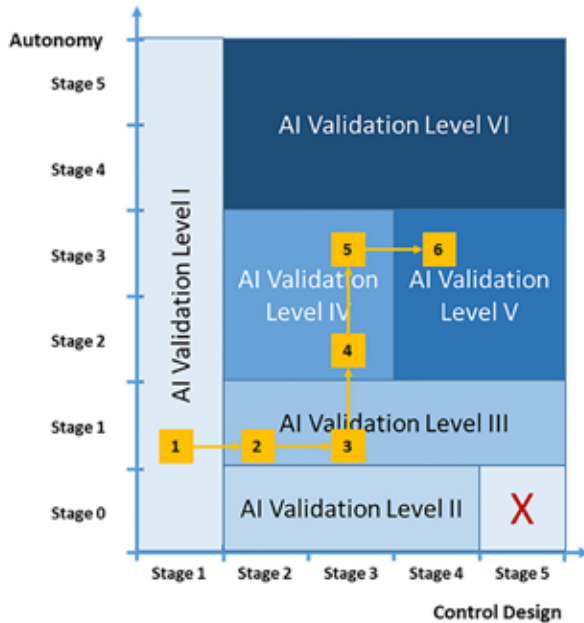
Systems in AI validation level VI are self-learning systems. It is expected that in the near future, strategies will be available for the control of continuous learning systems. There is no validation concept available now to ensure regulatory compliance for systems in this category.

In summary, the framework describes a tradeoff between the organizational burden of controlling the AI system during operation, which is more pronounced at the lower levels of the framework, and the technical requirements that facilitate increased validation activities to secure an increasingly autonomous AI system (see Figure 2).

## MATURITY POSITIONING AND DYNAMIC PATH

By following the framework outlined previously, the control design of an AI system is supported with regards to the following facets:

**Figure 3:** Example of a dynamic AI system's path within the autonomy and control framework.



A decision has to be made about how much human control should be embedded into the operative process.

**Initial design**

During the initial design of the AI control mode, based on critical thinking, a decision has to be made about how much human control should be embedded into the operative process. For instance, for a first start, a mode might be chosen with less autonomy and more control, hence reducing the requirements with regard to the technical framework, yet with a higher operational burden. This decision should be critically founded on the intention of use and the risk portfolio of the specific AI system and on the company's experience with AI system design and maintenance in general. The risk assessment mechanics specific to AI are not addressed in this article.

**Dynamic path**

Once the AI system has been established, it should be continuously evaluated for whether the control design and the positioning in

the maturity space are still appropriate, considering results from validation activities, post-market monitoring, and risk assessment updates, and from a business point of view, the balance of operational and technical burden. This evaluation may direct the design in either direction, e.g., the control design may be tightened (with more human control, less autonomy), given newly identified risks, or the AI system's autonomy may be expanded, accompanied by tighter technical control measures.

Management may consider the maturity model as a strategic instrument in order to dynamically drive the AI solution through its life cycle with regard to the system's autonomy and human control.

Example (see Figure 3):

1. The corporation decides to explore the usability of an AI system for a specific use case, parallel to an existing GxP-relevant process (AI validation level I).
2. After successful introduction of the AI system, the AI system should take over the GxP process, while still in a locked-state operation mode and controlled for all instances (AI validation level III); at the same time, stricter technical and functional validation activities are introduced.
3. Extending the AI system's value-add further, the control design is changed to a mode in which not all instances are controlled; because it's still operating in locked-state mode, AI validation level III applies. However, further controls such as sample checks of instances may be introduced, given the criticality of the GxP process.
4. After having collected sufficient experience with regard to the AI system in its specific use case, the autonomy is increased such that the system may indicate necessary retraining (AI validation level IV).
5. Extending the autonomy of the system further, the training process is now more oriented to the AI system's mechanics, i.e., in the way the retraining is performed, but the activation of such a new version is still verified by a human operator (still AI validation level IV).
6. As the final step in the solution's growth path, the control stage 4 is chosen so that the system controls itself (AI validation level V).

**CONCLUSION**

Because of the lack of AI/ML-specific regulations, other ways to determine the appropriate number of validation activities are required by the industry and development partners. The AI maturity model described in this article provides the rationale for the distinction among validation levels based on the AI model's stage of autonomy and control design. We consider this maturity model as the starting point for further discussions and as the basis for a comprehensive guideline for the validation of applications based on AI/ML in the pharmaceutical industry. We believe that our model has great potential for application in other life sciences industries. 🌀

## References

1. Fotheringham, S. "Technology Trends: The Transition to Digitalization." *Pharmaceutical Engineering* 42, no. 5 (2020). <https://ispe.org/pharmaceutical-engineering/september-october-2020/technology-trends-transition-digitalization>
2. International Society for Pharmaceutical Engineering. *GAMP® RDI Good Practice Guide: Data Integrity by Design*. North Bethesda, MD: International Society for Pharmaceutical Engineering, 2020.
3. Deutsches Institut für Normung e. V. "German Standardization Roadmap on Artificial Intelligence." November 2020. <https://www.din.de/en/innovation-and-research/artificial-intelligence>
4. US Food and Drug Administration. "FDA Discussion Paper: Artificial Intelligence and Machine Learning in Software as a Medical Device." 22 September 2021. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>

## Acknowledgments

The authors thank the ISPE GAMP D/A/CH Working Group on AI Validation, led by Robert Hahnrahts and Joerg Stueben, for their assistance in the preparation of this article and Martin Heitmann and Mathias Zahn for manuscript review.

## About the Authors

**Nico Erdmann** is a Manager at Deloitte Germany with a focus on regulatory and compliance in the life sciences industries. As part of his specialization in GxP compliance and IT quality, he supports life sciences companies in the digitalization of GMP processes and their validation, implementing smart labs, optimizing production process, and introducing continuous process verification concepts and leading all computerized system validation activities.

Nico developed a strong technological focus on implementing artificial intelligence in the GxP-regulated environment and developed validation concepts for their quality assurance.

**Rolf Blumenthal** is a Senior Consultant at Werum and has many years of experience in the pharmaceutical industry. He holds a degree in mathematics and computer science from the University of Dortmund, Germany. He joined Werum in 1979 and is one of the 15 employees who took over the company after the death of the founder in 1982. Initially, he gained expertise in the management of large MES software projects. In the following years, Rolf headed the development of Werum's MES product, established a consulting division, and started a new division to unlock the added value of data from the pharmaceutical industry. He is a recognized expert with extensive knowledge and hands-on experience in cGMP, FDA requirements, GAMP® guidelines, and computer validation. He is a member of the GAMP D/A/CH Steering Committee and the AI Validation Working Group.

**Ingo Baumann** is Head of Delivery and a Partner at Thescon. He holds a diploma in chemical engineering from the European University of Applied Sciences Fresenius. Prior to Thescon, Ingo held positions in pharma units at IBS (now Siemens Industry Software) and IDS Scheer. With more than 20 years of experience in the GxP-regulated industries, Ingo is an expert on computerized systems validation, data integrity, and internal control systems. He is a senior advisor for regulated master data and product life cycle management. As a member of the ISPE GAMP Special Interest Group for Software Automation and Artificial Intelligence and GAMP D/A/CH's working group for AI in GxP areas, he contributes to initiatives to establish guidelines and best practice approaches for the validation of artificial intelligence and machine learning applications. He joined ISPE in 2020.

**Markus Kaufmann** is Certified Assessor for Software Process Improvement and Capability Determination (SPICE) and is working as Global GMP Auditor at Novartis in Basel, Switzerland. He holds a degree in process engineering from the University Karlsruhe, Germany. Since 1998, Markus has worked in a GMP-regulated environment at pharmaceutical companies including Pfizer, Roche, and Bayer in various positions from development to distribution, automation, and IT.

# Please Join Us in Thanking ISPE's 2022 Corporate Partners



Through the ISPE Corporate Partnership program, these companies have committed to supporting and contributing to ISPE's mission within the pharmaceutical industry.

