



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Good manufacturing practice (GMP)

Multistakeholder workshop on expert contributions to artificial intelligence guidance development (Annex 22)

30 June 2026



Table of contents

Introduction 3

1. Topic Discussions 4

1.1. Topic 1: Regulatory Pathways for Adaptive/probabilistic AI models in GMP 4

1.2. Topic 5: Strategic Risk & Compliance Limits.....5

1.3. Topic 3: Human Oversight & Accountability 7

1.4. Topic 2: Technical Reliability & Incident Response 8

1.5. Topic 4: Validation & Lifecycle Management..... 10

1.6. Topic 6: Cybersecurity and Outsourced activities 12

2. Conclusion. 13

3. Next Steps 14

Introduction

Annex 22 “Artificial Intelligence” will be an Annex of the EU Good Manufacturing Practices (GMP) Guide intended to provide additional guidance to Annex 11 for computerised systems in which AI models are embedded.

The draft version of Annex 22 currently states “the document does not apply to Generative AI and Large Language Models (LLM), models with a probabilistic output, dynamic models and models which adapt performance, such models should not be used in critical GMP applications.”

Comments received during the public consultation on Annex 22 indicated support for allowing manufacturers to use Large Language Models and generative AI systems in GMP applications, whether critical or non-critical. This position reflects the need to support innovation and encourage investment by pharmaceutical companies.

The drafting group has therefore discussed the scope of Annex 22 to allow dynamic or adaptive models, probabilistic models, and Generative AI/LLMs in GMP applications, provided they comply with the requirements of the annex and are supported by a fully documented, robust, risk-based control strategy.

However, concerns remain about the elements needed for such a control strategy. The drafting group therefore requested a meeting with AI experts to assess whether a risk-based approach can be applied to Generative AI/LLMs, including the use control mechanisms as mitigations for their use in GMP applications.

The objective of this workshop was therefore to gather structured, experience-based input from industry on the control, governance, and mitigation measures applied throughout the AI system lifecycle, with particular emphasis on the application of ICH Q9 quality risk management principles. To facilitate this discussion, industry stakeholder associations registered with European Medicines Agency (EMA) collaborated to provide a consolidated response to the List of Questions previously shared with industry and delivered during the workshop a series of presentations reflecting a collective industry perspective across the six predefined topics.

These 6 topics addressed key areas including:

Topic 1: Regulatory Pathways for Adaptive/probabilistic AI models in GMP

Topic 2: Technical Reliability & Incident Response

Topic 3: Human Oversight & Accountability

Topic 4: Validation & Lifecycle Management

Topic 5: Strategic Risk & Compliance Limits

Topic 6: Cybersecurity and Outsourced activities

Through the presentations and subsequent discussions, the workshop was intended to enhance the drafting group's understanding of current practices, challenges, and emerging approaches for managing risks associated with AI systems in GMP environments. The insights obtained were intended to inform the revision of the scope of Annex 22 and support the development of a risk-proportionate and implementable regulatory framework that both safeguards product quality and patient safety while enabling responsible innovation.

1. Topic Discussions

1.1. Topic 1: Regulatory Pathways for Adaptive/probabilistic AI models in GMP

Questions from the Drafting Group

- How could we accommodate adaptive and probabilistic models in Annex 22?
- What would be the validation paradigm for adaptive models that should be in annex 22?
- Please comment on the application of a quality risk management approach to the entire lifecycle of such AI system models that may be used in GMP high risk areas where there may be a low detectability of deficiencies but a high impact on the patient and taking into account ICH Q9R1 principles. In applying quality risk management, what should the regulated user take into account when selecting and using an adaptive or probabilistic model?

Industry Perspective

The industry position, presented as a consolidated interested parties view, was that adaptive, probabilistic and generative AI models should not be excluded from GMP applications on the basis of a technology label alone. Instead, the central criterion should be whether the model can be demonstrated to be fit for its intended use and whether the associated risks can be controlled within an existing pharmaceutical quality system. The concepts of importance, uncertainty, complexity and level of formality should be taken into consideration when a risk assessment is conducted. The speaker emphasised that AI is a new and rapidly evolving technology, but that uncertainty and model drift are not conceptually new to GMP. Measurement systems drift, calibration intervals are adjusted on a risk basis, and lifecycle review is already used to confirm that a system remains fit for use. This analogy was not intended to equate all forms of AI adaptation with measurement drift, but to illustrate that GMP already contains mechanisms for reassessing continued suitability when system behaviour changes over time.

A key theme was the distinction between uncertainty and unacceptable risk. The presentation argued that uncertainty should drive increased knowledge generation, stronger lifecycle controls, more formal validation, monitoring and re-verification, rather than an automatic prohibition. The proposed Annex 22 wording presented by the speaker would broaden the scope to cover deterministic and probabilistic outputs, static and dynamic models, and models whose performance may adapt during use, provided controls are implemented according to intended use and risk to patient safety and product quality.

Q&A session

The Q&A focused on whether the framework sufficiently addresses specific features of generative AI and LLMs, including hallucination, fabrication, sycophancy or epistemic overconfidence. Regulators questioned whether these phenomena are merely ordinary statistical uncertainty or a fundamentally different failure mode requiring specific attention. The response was that such behaviours must be identified and evaluated during development and validation; if the manufacturer cannot demonstrate fitness for intended use in view of those failure modes, the use cannot be justified. The discussion did not set a specific technical method for estimating confidence intervals or hallucination rates; rather, it deferred detailed technical controls to later sessions.

Detectability was another point for discussion. The moderator noted that in adaptive or generative AI, inability to detect failure may itself be a central risk factor. The speaker acknowledged that, although

the general definition of risk in ICH Q9 refers to severity and probability, methods such as Failure Mode and Effects Analysis (FMEA) may use detectability because it is useful in certain contexts. For AI, detectability may be highly appropriate as a dimension of the assessment, particularly where errors are difficult to observe. The discussion therefore supported consideration of detectability, without mandating a single risk scoring method.

The session also touched on complexity. One regulator asked whether inclusion of complexity might be controversial for industry. The speaker responded that the interested parties involved in preparing the workshop had not objected; on the contrary, complexity is explicitly part of the ICH Q9(R1) formality concept and should be included when determining the necessary rigour of controls. Overall, the session created broad consensus that Annex 22 should remain technology-neutral and principle-based, but should make clear that AI models require documented intended use, documented understanding of uncertainties, lifecycle monitoring, risk review, and proportionate control before use in GMP applications.

Key Messages from the Session

- Existing QRM and lifecycle principles are applicable to AI, including adaptive and probabilistic models.
- Technology labels should not determine acceptability; intended use, impact, uncertainty, complexity and control strategy should.
- Higher risk should drive increased formality and stronger controls.
- Detectability is a relevant AI risk consideration, especially where deficiencies are difficult to identify.
- Annex 22 should avoid detailed technical prescriptions and remain coherent with Annex 11, Chapter 4 and ICH guidance.

Potential Impact on Annex 22 Drafting

The discussion explored the considerations relevant to the possible use of probabilistic or adaptive models, including defined intended use, risk assessment, documented knowledge of the model and data, lifecycle validation, monitoring and risk-based change control. Definitions may need to clarify intended use, model influence, complexity, uncertainty, control strategy, detectability and continued fitness for use.

- Industry would be expected to justify use through documented evidence and lifecycle controls.

1.2. Topic 5: Strategic Risk & Compliance Limits

Questions from the Drafting Group

- Do current guardrail approaches sufficiently address GMP high risk areas such as data integrity, process control, auditability, and traceability—especially where GenAI systems may influence or automate decision-making?
- Where do experts believe the limits of risk-based mitigation lie? Are there classes of critical decisions where no combination of guardrails and oversight would be sufficient?
- Are there scenarios in which the level of guardrail effort required to mitigate GenAI risks becomes disproportionate or operationally impractical, thereby indicating that such systems should not be used in certain critical GMP functions?

- What other elements of an overall control strategy should be considered for inclusion in the Annex 22 to allow the use of dynamic or adaptive models, probabilistic models, and Generative AI/LLMs in GMP applications?

Industry Perspective

The second session addressed whether there are strategic limits to the use of generative AI or other probabilistic/adaptive AI models in critical GMP decision-making. The interested parties' position was that guardrails and other controls, when designed under QRM principles, can contribute to a state of control for data integrity, traceability, auditability and process control. The presentation argued that these attributes are not new GMP expectations; industry has long managed them through systems, records, audit trails, verification and monitoring. However, AI requires that established controls and newer AI-specific controls operate together as an integrated framework.

The industry position was strongly against categorical prohibitions. The speaker stated that there should be no class of decisions considered unacceptable solely by technology category or criticality label. Instead, acceptability should be assessed case by case by considering decision consequence, model influence, uncertainty, complexity and the ability to reduce residual risk to an acceptable level. The presentation acknowledged that in some use cases the mitigation effort may be heavy, impractical, or insufficient. In such cases the conclusion would be that the use is not justified at that time and in that context, and the conclusion should be documented.

Q&A session

A significant regulatory challenge raised during Q&A was how a manufacturer could demonstrate to inspectors that the combined layered controls do provide reliable output for a critical decision. The response identified the risk assessment as the starting point: it should describe the risk scenarios, the controls mapped to those scenarios, and the evidence generated through verification and monitoring. Monitoring history, responsiveness to signals, and documented state of control would then be inspection evidence. Additional discussion introduced the concept of controls by design, continuous monitoring of expected residual risk, and automated controls where feasible.

Responsibility and accountability were also explored. Regulators asked how a responsible person could take responsibility for a model that is not fully transparent. Responses emphasised that the use of AI should not remove predicate-rule responsibilities and that accountable persons may rely on qualified experts, as they already do for complex manufacturing technologies. The idea of an "AI qualification dossier" was introduced by another industry speaker as a package containing validation evidence, training data information, guardrails around prompting and output, and evidence that the system works as expected. Overall, the strategic session reinforced the preference for Annex 22 as an outcome-based framework, not a technology-prescriptive checklist.

Key Messages Emerging from the Discussion

- No categorical prohibition was supported by industry; use and risk should be assessed in context.
- A valid QRM outcome may still be 'do not use' where uncertainty cannot be controlled.
- State of control should be demonstrated through mapped risk scenarios, controls, verification and monitoring evidence.
- Criticality alone is insufficient; model influence, uncertainty, complexity and detectability must be considered.
- Accountability remains with qualified/responsible persons and cannot be delegated to AI.

Potential Impact on Annex 22 Drafting

Annex 22 may be influenced towards an explicit performance-based acceptability framework: identify intended use; evaluate decision consequence, model influence, uncertainty, complexity and detectability; define controls; verify them; monitor residual risk; and document the go/no-go conclusion. Definitions of direct impact, model influence, control strategy and residual-risk acceptance may need strengthening.

- Industry expectations would include an inspectable evidence package.

1.3. Topic 3: Human Oversight & Accountability

Questions from the Drafting Group

- What level and form of human-in-the-loop oversight is still required when guardrails are implemented, and is this oversight sufficient to ensure accuracy, traceability, and accountability?

Industry Perspective

The third session examined the role of human oversight when technical guardrails and AI-supported workflows are used. The interested parties emphasised that human oversight should be designed through a documented QRM process and should be proportionate to the AI model's risk profile and context of use. A key distinction was made between "human oversight" as a broad lifecycle governance objective and "human in the loop" as one specific implementation pattern. Human oversight asks whether a qualified, accountable human can understand, challenge, intervene, override, reject or escalate. Human-in-the-loop, by contrast, is where the process waits for a human action before a GMP-relevant workflow proceeds.

The presentation placed human oversight within an integrated control framework, not as a standalone safeguard. Guardrails may reduce risk but do not eliminate the need for risk-proportionate human oversight or accountability. EU GMP Part I Chapter 2 on personnel qualification and training was cited as applicable, and the EU AI Act was referenced in relation to human oversight commensurate with risk, autonomy and context of use. The oversight model proposed criteria such as impact of incorrect decisions, model influence, uncertainty, complexity and importance. These criteria determine the degree of human involvement and the permissible level of autonomy.

A deviation management support use case illustrated how human involvement might be high when a model has high impact, variable uncertainty and complexity, even if the AI influence is initially more informational. Over time, as knowledge and experience increase, the risk profile may shift, potentially allowing more autonomy and less human involvement. Conversely, if new risks emerge, oversight may increase. This reinforced the lifecycle nature of oversight.

Q&A session

The Q&A focused on whether human oversight can be meaningful in practice. Regulators asked how to monitor the performance of the human performing oversight; how to manage automation bias; what information and competence a human reviewer needs; whether human-in-the-loop can become a bottleneck; how fallback/override should be validated; and how oversight can be evidenced during inspection. Responses identified training, qualification, knowledge testing, audit trails, recording of AI outputs and human modifications, and monitoring of human-AI team performance as potential evidence. Automation bias was acknowledged as requiring meaningful oversight, appropriate training

and technical measures. A human reviewer does not need to understand every internal detail of the model, but must understand the relevant model principles, the process, the output and the decision context sufficiently to challenge the AI.

A regulator asked if human-in-the-loop could ever justify reduced validation. The question was deferred to the validation topic; an industry speaker noted that the answer is risk-based rather than a categorical yes or no. Separately, one speaker noted that oversight alone does not directly correct risk; it provides knowledge, detection and potential triggers for action. Therefore, human oversight should be judged by whether it produces relevant knowledge and supports effective risk reduction through follow-up actions. The discussion also recognised that human reviewers may face operational pressure or fatigue. The response was that such risks are not AI-specific and should be managed through existing GMP provisions. Overall, the session supports Annex 22 language requiring meaningful, risk-based and documented oversight rather than default, symbolic or one-size-fits-all human-in-the-loop.

Key Messages Emerging from the Discussion

- Human oversight is broader than human-in-the-loop and should be lifecycle-wide.
- Human-in-the-loop is one possible control, not a default requirement for all AI uses.
- Oversight must be proportionate to risk, autonomy, uncertainty, complexity and decision impact.
- Training, qualification, authority and ability to challenge are essential.
- Oversight should be evidenced and periodically reviewed; it cannot compensate for inadequate validation.

Potential Impact on Annex 22 Drafting

The discussion supports Annex 22 text requiring human oversight to be defined in the control strategy according to intended use and QRM. Definitions should distinguish oversight, Human in the Loop (HITL), Human in the Loop (HOTL) monitoring, fallback/override and accountability. Requirements may include training, qualification, documentation, auditability and periodic review of oversight effectiveness.

- Industry expectations should include evidence that reviewers can understand, challenge and intervene.

1.4. Topic 2: Technical Reliability & Incident Response

Questions from the Drafting Group

- To what extent can available guardrail mechanisms reliably prevent, detect, or contain hallucinations, incorrect recommendations, or fabricated data when dynamic or adaptive models, probabilistic models, and Generative AI/LLMs are used within GMP workflows?
- When guardrails fail or detect uncertainty, what mechanisms are required to ensure timely escalation and prevention of GMP impact?

The fourth session provided the most technical discussion of the day. It examined whether available guardrail mechanisms can prevent, detect or contain hallucinations, incorrect recommendations and fabricated data, and how failures should be escalated. The presentation used five use cases:

- adaptive Heating, Ventilation and Air Conditioning (HVAC) control within validated environmental boundaries;
- an agentic AI system supporting deviation management;
- an LLM supporting GMP risk assessment for an electronic batch record system;
- text-to-SQL generation for annual product review data extraction;
- and model predictive control for bioreactor feeding.

Some examples were stated to be pilots or theoretical frameworks rather than production systems.

A central model was layered control:

- input/prevention controls,
- in-model/detection controls,
- and output/containment controls.

This was linked to the “Swiss cheese” concept: each control addresses a particular failure mode and may be imperfect alone, but layers can combine to reduce risk.

Examples of input controls included context-of-use verification, out-of-distribution detection, semantic validation and rule-based limits. In-model controls included retrieval augmented generation with source/citation binding, conformal prediction, uncertainty quantification, self-consistency scoring and access to model logs or rationale where available. Output controls included hard constraints, confidence gates, cumulative sum monitoring, exponentially weighted moving averages, ensemble disagreement, human gates and fallback to validated states.

The presenter repeatedly emphasised that guardrails are specific controls for specific risks and that unanticipated failure modes cannot be assumed to be covered by controls designed for known failure modes.

This led to important Q&A. One regulator asked whether all controls must be derived from risk assessment, given that GMP often uses tacit knowledge and experience to create controls beyond formally identified risks. The response, supplemented by another speaker, was that risk assessments are limited by available knowledge and must be updated through lifecycle learning, deviations and defects. New knowledge should feed back into risk assessments and therefore into new or revised controls.

Terminology was a major issue. Several participants questioned whether the term “guardrail” should appear in Annex 22. Industry responses suggested that “control mechanisms” may be preferable in the main annex because guardrails can be technology-specific and transient. The discussion reflects agreement that the main annex should contain principles and objectives — such as prevention, detection and containment — while technical methods and use cases may be better placed in appendices, Q&A, or another updateable format.

The Q&A also explored whether these methods imply a preference for local or self-developed models. The presenter stated that the key is not ownership of the model but access to sufficient evidence to measure how it works. If a third-party model or Application Programming Interface (API) cannot provide necessary outputs, rationale, logs or evidence needed for control, the manufacturer cannot measure and justify the system adequately. The session also addressed fallback. For HVAC and bioreactor examples, failure or out-of-bound behaviour led to reversion to a pre-validated classical control state. Regulators noted that a fully model-operated process might not have such a classical fallback. The response was that this would need to be considered in the contingency strategy;

sometimes accepting stoppage may be the contingency. Overall, this session strongly supports Annex 22 requiring layered, risk-based controls and documented escalation pathways, while keeping technical examples outside the main body or at least non-prescriptive.

Key Messages Emerging from the Discussion

- Controls should be layered across prevention, detection and containment.
- Each control should be linked to intended use, failure modes and measurable acceptance criteria.
- Unanticipated failures require lifecycle learning and risk assessment updates.
- The main annex should avoid transient technical details; examples may be placed in updateable guidance.
- Supplier or API models must provide sufficient evidence and transparency for control.

Potential Impact on Annex 22 Drafting

This discussion may significantly shape Annex 22's control-strategy sections. The annex could require layered controls and escalation pathways while avoiding detailed references to any one technical method. Definitions may need to cover control mechanisms, unacceptable output, failure mode, escalation, fallback, safe state and monitoring threshold.

- Industry expectations include documented qualification of controls, evidence of thresholds and lifecycle monitoring.

1.5. Topic 4: Validation & Lifecycle Management

Questions from the Drafting Group

- How can guardrails remain effective when the underlying AI model evolves (e.g., updates, retraining, drift), and what controls are necessary to ensure guardrail performance remains validated over time?
- What type and level of evidence (e.g., validation data, stress testing results, failure analyses) would be needed to justify that guardrails are a reliable risk mitigation measure enabling GenAI use in GMP functions?
- Can you demonstrate the effectiveness of such guardrails or similar measures?

Industry Perspective

The fifth session addressed validation and lifecycle management as the integrative framework connecting the other topics. The presenters framed validation lifecycle management as central to the six-pillar architecture because it interfaces with regulatory pathways, technical reliability, human oversight, change control, monitoring and incident response. The session answered questions on how controls remain effective when models evolve, what evidence is needed to justify control mechanisms as reliable risk mitigations, and how the effectiveness of such controls can be demonstrated.

The presentation emphasised a lifecycle decision-making framework rather than a single validation event. A model's intended use, design assumptions, context, data, training and verification strategy should define the depth and type of validation. Risk factors discussed included: uncertainty,

importance, complexity, novelty of the AI approach, supplier experience, model limitations, decision consequence, interpretability/explainability, degree of autonomy and adaptiveness.

These factors should determine validation activities, selection of control mechanisms, evidence requirements, monitoring and revalidation/change triggers.

The presenters highlighted that data used for validation/testing should remain independent from development and training data.

Q&A session

During Q&A, regulators asked how independence can be maintained where a model continuously re-trains on operational data. The answer was to use lifecycle mechanisms, defined samples, test sets, procedures and controls to preserve a means of independent verification; where the model evolves, monitoring and change-control triggers should determine when additional verification or re-optimisation is required. The response was not fully prescriptive and did not define one mandatory technical approach.

The session also dealt with human-in-the-loop, fallback and oversight in the validation context. Discussion from earlier sessions was carried forward: human oversight cannot be a substitute for inadequate validation, and control effectiveness must be evidenced. There was recognition that automation bias, human bias, skill level, AI literacy, explainability, confidence scores and challenge/testing of operators may be relevant. Lifecycle monitoring should include incidents and problems, Corrective and Preventive Action (CAPA), rollback or intervention criteria, supplier oversight and cloud/provider change visibility. The presenters linked supplier oversight to validation because the regulated user's responsibility cannot be transferred to the supplier.

A recurring topic was use-case patterns. Presenters suggested that risk patterns may be developed for similar AI use-case types, including risks such as cybersecurity, model evasion or data poisoning, and that applicable control strategies could be mapped accordingly. However, not all controls fit every use case. The workshop did not conclude a definitive template, but the concept of reusable patterns emerged as an idea for further discussion.

Overall, this topic reinforced that Annex 22 should require a documented AI lifecycle: intended use, design/qualification, independent verification, acceptance criteria, monitoring, change control, revalidation/re-verification, incident and problem management, CAPA, supplier oversight, and periodic risk review. It also reinforced the need to keep technical examples flexible and updateable.

Key Messages Emerging from the Discussion

- Validation should be a lifecycle, not a one-time event.
- Validation depth should be driven by intended use, uncertainty, importance, complexity, autonomy and adaptiveness.
- Independent verification evidence remains a critical expectation, but implementation for continuously learning systems needs clarification.
- Monitoring, change control, incident/problem management and CAPA are part of validation lifecycle management.
- Supplier and cloud-provider oversight are integral because accountability remains with the regulated user.

Potential Impact on Annex 22 Drafting

The validation session may influence the guideline text on lifecycle control, continued model performance, revalidation/re-verification and change management. Definitions may need to include lifecycle validation, independent test set, operational monitoring, model drift, re-optimisation, change trigger and control effectiveness.

- Industry expectations would include an AI validation/qualification package aligned to risk and intended use.

1.6. Topic 6: Cybersecurity and Outsourced activities

Questions from the Drafting Group

- Is there a need to address prevention and protection of AI systems from tampering or unauthorized access in Annex 22 or is this already sufficiently addressed in Annex 11? What would you propose in terms of appropriate wording?
- What risks arise when guardrail infrastructure is outside the manufacturer's direct control, and how can supplier qualification, change-control visibility, and independent audit capabilities be preserved in such cloud-based AI supply chains?

Industry Perspective

The final topic addressed cybersecurity and outsourced activities. The presenter began from the premise that AI systems consist of data, model/software and hardware; therefore, from a GMP perspective they remain computerised systems. On this basis, the interested parties' position was that existing Annex 11 principles, particularly Chapter 7 on outsourced activities, already provide an appropriate framework for cybersecurity and supplier management, including for AI systems. The presentation also referenced broader cybersecurity frameworks and standards such as NIS2, SOC 2 and ISO 27000 series, and noted that the EU AI Act contains extensive cybersecurity references.

The first core message was that AI does not require a separate cybersecurity regime in Annex 22 if Annex 11 and applicable cybersecurity standards are properly applied. Access controls, audit trails, supplier qualification, continuous monitoring, agreements and legal arrangements were described as already covered by Annex 11. The second message was that the regulated pharmaceutical user remains accountable for the computerised/AI system even when a third party provides the model, platform or cloud infrastructure. Responsibilities may be shared operationally, but accountability for GMP use remains with the regulated user.

Supplier and cloud oversight were discussed in detail. The presenter argued that agreements must clarify responsibilities, configuration expectations, change control participation, evidence access, audit/certification arrangements and transparency of controls. The supplier must understand pharmaceutical requirements and language. For cloud and AI services, the pharmaceutical user should require evidence for relevant controls, data integrity, anonymisation/minimisation, zero data retention where applicable, confidential computing or trusted execution environments, regional requirements, federated learning and data loss prevention where relevant. The presentation framed this as a shared-responsibility model in which the provider and customer responsibilities vary depending on infrastructure-as-a-service, platform-as-a-service or software-as-a-service models.

Q&A session

The Q&A focused on what objective assurance evidence is needed where direct physical audit of major cloud suppliers is difficult. The answer was to use documented evidence such as certifications and compliance information, including NIS2, SOC 2, ISO 27000 series and European framework suitability, and then assess alignment with the company's own requirements. If the provider cannot meet requirements, another option should be chosen. A chat question asked whether current GMP requirements from Annex 11 Chapter 7 are sufficient for AI suppliers/service providers. The response was that Annex 11 Chapter 7 is fully applicable because AI systems are computerised systems; however, agreement types may vary depending on the service and intended use.

The discussion also linked back to previous topics: where there is limited visibility of supplier changes, regulated users need to be part of change control and have visibility over control logic. Control of outsourced AI therefore links to validation lifecycle and monitoring. The final session supports an Annex 22 approach that references Annex 11 rather than duplicating it, while potentially adding AI-specific considerations around evidence, transparency, change visibility, data retention, confidential computing and shared responsibilities.

Key Messages Emerging from the Discussion

- AI systems should be treated as computerised systems for cybersecurity and outsourcing governance.
- Annex 11 Chapter 7 was considered applicable and sufficient at a principle level.
- The regulated user remains accountable even when using third-party AI or cloud services.
- Supplier agreements must address responsibilities, evidence, configuration and change control.
- AI-specific examples may be useful, but duplication of Annex 11 should be avoided.

Potential Impact on Annex 22 Drafting

Annex 22 may be influenced to cross-reference Annex 11 for cybersecurity and outsourced activities while adding targeted AI-specific considerations. Definitions may include shared responsibility, third-party AI provider, regulated user accountability, supplier evidence and change visibility.

- Industry expectations would include supplier qualification, agreements, certifications/evidence, documented responsibilities and continuous monitoring.

2. Conclusion.

The workshop provided valuable expert input to support the ongoing development of Annex 22 and demonstrated that a broad range of industry stakeholders were able to articulate a largely aligned position across the six discussion topics. While recognising the challenges associated with adaptive, probabilistic and generative AI systems, participants consistently advocated for a risk-based, technology-neutral approach in which the acceptability of an AI application is determined by its intended use, the risks it presents, and the effectiveness of the controls applied throughout its lifecycle rather than by the underlying technology itself.

Across all sessions, several common themes emerged. These included the importance of documented intended use, robust quality risk management, lifecycle validation and monitoring, risk-proportionate

human oversight, effective control mechanisms, clear accountability, supplier oversight, and the maintenance of a demonstrable state of control. Participants also emphasised that uncertainty, complexity, model influence and detectability are important considerations when evaluating AI systems used in GMP-regulated activities.

The discussions further demonstrated that existing GMP principles, particularly those relating to quality risk management, validation, lifecycle management, oversight, and outsourced activities, remain relevant and applicable to AI systems. At the same time, the workshop helped identify areas where additional clarification, definitions, or AI-specific expectations may be needed to ensure that Annex 22 provides a practical and scientifically sound framework for the use of increasingly sophisticated AI technologies in GMP environments.

Overall, the workshop achieved its objective of gathering structured, experience-based input from industry on the governance, control, validation and oversight of AI systems and provided the drafting group with valuable insights to support the revision of Annex 22. The discussions suggest that subject to appropriate controls and a robust risk-based framework, there may be viable pathways for the use of adaptive, probabilistic and generative AI technologies in GMP-regulated applications while continuing to protect product quality and patient safety.

3. Next Steps

The information presented during the workshop, including the consolidated responses provided by the participating industry associations, will be carefully reviewed and considered by the Annex 22 drafting group.

Following the workshop, the drafting group will assess the key themes, common approaches, areas of divergence, and identified challenges discussed across the six topics. The outcomes of these discussions will be used to inform the ongoing revision of Annex 22, with the aim of developing a science-based, risk-proportionate, and implementable framework for the use of AI in GMP-regulated environments. Where necessary, additional clarification and internal discussion may be undertaken to ensure that the revised guidance appropriately balances the protection of product quality and patient safety with the potential benefits of technological innovation.

European Medicines Agency

Domenico Scarlattilaan 6
1083 HS Amsterdam
The Netherlands
Telephone +31 (0)88 781 6000

www.ema.europa.eu

© European Medicines Agency, 2026

Reproduction is authorised provided the source is acknowledged. For any use or reproduction of images or other materials that are not under the copyright of the European Medicines Agency as specified subsequently, permission must be sought directly from the respective copyright holders. The indicated images or other materials are copyright protected and have been licensed by the European Medicines Agency via Adobe Stock, Dreamstime or iStockphoto, as shown respectively in the credit lines of each image or other material.

Cover image: © Lee | stock.adobe.com. AI generated with AI

