

WHITE PAPER NO. 12

Signal Detection in the Age of Artificial Intelligence

From Disproportionality Screening to Human-Governed,
AI-Augmented Signal Management

Reinhold J. Schilling

Version 1.0

September 2026

This white paper reflects the author's professional perspective on the use of artificial intelligence in pharmacovigilance signal detection and signal management. It is intended for educational and discussion purposes only and does not constitute legal, regulatory, medical or compliance advice.

AI-Augmented Signal Management: From Data to Accountable Action

1.

SAFETY DATA SOURCES

Cases, literature, studies, registries, complaints and regulatory data.

2.

AI-SUPPORTED PREPARATION

Retrieve, normalise, deduplicate, extract and structure.

3.

PATTERN RECOGNITION

Screen associations, cluster evidence and prioritise review.

4.

ALERT OR HYPOTHESIS

An output requiring qualified review, not a signal conclusion.

5.

HUMAN ASSESSMENT

Interpret evidence, validate and document scientific judgement.

6.

ACCOUNTABLE ACTION

Determine benefit-risk implications and regulatory measures.

AN AI-GENERATED ALERT IS NOT A VALIDATED SIGNAL.

RETRIEVE AND STRUCTURE

Locate evidence, extract defined information, normalise terminology and organise controlled source data.

DETECT AND PRIORITISE

Identify patterns, rank alerts, cluster cases and support the allocation of expert review.

SUMMARISE AND ORCHESTRATE

Prepare traceable drafts, organise evidence, route tasks and monitor workflow status.

NON-DELEGABLE HUMAN RESPONSIBILITIES

Clinical interpretation • Signal validation and scientific assessment • Regulatory and benefit-risk decisions

GOVERNANCE FOUNDATION

Intended Use | Data Suitability | Scientific Validation | Human Oversight | Auditability | Lifecycle Monitoring | Resilience and Fallback

Success is measured at the level of the complete human-machine workflow, not by model accuracy alone.

Disclaimer

This white paper reflects the author's professional perspective on the use of artificial intelligence in pharmacovigilance signal detection and signal management. It is intended for educational and discussion purposes only and does not constitute legal, regulatory, medical or compliance advice.

Organisations remain responsible for determining the applicability of legal and regulatory requirements to their products, systems, processes and jurisdictions. References to technologies, methodological approaches or governance practices do not constitute endorsement of a particular product or provider.

Legal classification under the European Union Artificial Intelligence Act and other applicable legislation should be assessed for each use case with appropriately qualified legal, regulatory, data-protection and technical specialists.

Executive Summary

Signal detection is entering a period of significant technological change. Pharmacovigilance organisations must evaluate increasing volumes of individual case safety reports, scientific literature, clinical and observational data, product complaints, medication errors, digital sources and other structured and unstructured information.

Artificial intelligence can support information retrieval, data structuring, pattern recognition, alert prioritisation and evidence synthesis. Machine learning may identify multivariable relationships that conventional pairwise screening does not readily reveal. Natural language processing can extract clinically relevant information from narratives and publications. Knowledge graphs can represent relationships between products, events, diseases and mechanisms. Generative AI can support retrieval, summarisation and drafting when its outputs remain grounded in controlled sources and subject to qualified verification.

AI also introduces material risks. These include false-positive and false-negative outputs, algorithmic bias, data leakage, overfitting, model drift, limited explainability, automation bias, alert fatigue, hallucinated content, cross-case contamination, prompt injection, privacy breaches and unclear accountability. Retrospective model performance does not establish prospective operational superiority, improved decision quality, earlier regulatory action or improved patient-safety outcomes.

Most importantly, an algorithmic alert is not automatically a safety signal, and a safety hypothesis is not automatically a validated signal. Signal management is a scientifically governed process comprising detection, validation, confirmation where applicable, analysis and prioritisation, assessment, recommendations for action, communication and tracking. Good Pharmacovigilance Practices Module IX remains an important methodological foundation,[1] while Commission Implementing Regulation (EU) 2025/1466[2] and current European Medicines Agency guidance[3,4] affect the operational interpretation of EudraVigilance monitoring and signal handling in the European Economic Area.

Traditional methods remain essential. Disproportionality screening, case-series assessment, medical review, literature evaluation and epidemiological investigation provide an established methodological foundation.[5] However, these methods can be challenged by the volume, velocity and heterogeneity of modern safety data.

The appropriate future model is therefore neither fully manual nor presumptively autonomous. It is a human-governed operating model in which validated technology performs tasks suited to computational analysis while named and qualified pharmacovigilance professionals retain authority for clinical interpretation, signal validation, signal assessment, benefit-risk conclusions, escalation, closure and regulatory action.

The appropriate unit of validation is the complete socio-technical workflow, not the model in isolation. This includes the data, model, software platform, interfaces, users, procedures, governance, documentation and resulting decisions. Model performance is relevant, but it is not sufficient evidence that the combined human-machine process is effective.

This white paper presents the **author-proposed HUMAN-SIGNAL Framework** as a practical structure for implementing AI-supported signal detection and signal management. It combines defined intended use, measurable performance, suitable data, scientific validation, accountable human oversight, inspection-ready documentation, lifecycle governance, ongoing monitoring, auditability and controlled change. Resilience and manual fallback are cross-cutting, non-negotiable principles.

The objective is not to prevent innovation. It is to ensure that innovation supports the primary purpose of pharmacovigilance: protecting patients through timely, scientifically defensible and appropriately documented safety decisions.

1. Introduction

Pharmacovigilance signal detection has always involved incomplete and uncertain information. A single well-documented case may raise a medically important concern. A statistical imbalance may indicate an association requiring investigation. A case series, publication, clinical study or observational analysis may materially change the understanding of a known risk.[6]

The challenge is not simply to find more associations. It is to identify information sufficiently credible and relevant to justify further investigation while avoiding the uncontrolled consumption of resources by weak, duplicated or clinically implausible alerts.

Potential sources of safety information include:

- Individual case safety reports
- Medical and scientific literature
- Clinical trial data
- Non-interventional studies
- Electronic healthcare data
- Disease and product registries
- Product complaints
- Medication errors
- Patient-support programmes
- Regulatory databases
- Digital and social media sources
- Molecular, pharmacological and toxicological information

As of February 2025, VigiBase held more than 40 million reports from over 180 members of the WHO Programme for International Drug Monitoring.[7] The database provides a global source of post-marketing information on adverse events reported in association with medicines and vaccines.

Greater data availability does not automatically improve signal detection. Spontaneous reporting is affected by underreporting, incomplete information, duplicate reporting, stimulated reporting, notoriety bias and the absence of reliable exposure denominators. Healthcare data may be affected by confounding, coding variability, selection effects and missing clinical context. Digital sources can contain repetition, uncertain provenance and unverifiable information.

AI is being introduced into this environment with the prospect of faster processing, better prioritisation and more consistent analytical support. These expectations require careful calibration. AI cannot remove limitations in the underlying data, convert a reporting association into an incidence estimate, transform correlation into causality or eliminate the need for medical judgement.

The central question is therefore not whether AI can generate alerts. It can. The more important question is whether a defined AI application can help pharmacovigilance professionals reach earlier, more consistent and more defensible safety decisions without introducing risks that are insufficiently understood or controlled.

2. Signal Detection and Signal Management

2.1 Distinct Concepts

The following terms should not be used interchangeably.

Algorithmic alert: A threshold exceedance, anomaly, statistical result or other machine-generated output indicating that predefined review criteria have been met.

AI-prioritised association: A product-event combination, case group or evidence item ranked by a model for review.

Safety hypothesis: A proposition that available information may indicate a new potentially causal association or a new aspect of a known association.

Potential signal: Information that has entered the signal-management process but has not completed validation.

Validated signal: A potential signal for which the available evidence and documentation support further analysis according to the applicable process.

Signal assessment conclusion: A scientific judgement reached after evaluating the relevant evidence, uncertainties, alternative explanations and clinical context.

Regulatory or benefit-risk conclusion: An accountable decision concerning communication, product information, additional pharmacovigilance, risk minimisation, authorisation conditions or another regulatory action.

An algorithmic alert may lead to a safety hypothesis. A safety hypothesis may enter validation. A validated signal may proceed to analysis and assessment. None of these steps predetermines the final regulatory or benefit-risk conclusion.

2.2 Signal Management

GVP Module IX describes signal management as a set of activities covering signal detection, validation, confirmation, analysis and prioritisation, assessment, recommendations for action and the tracking of resulting steps and decisions.[1]

Signal detection is therefore the beginning of a process, not its conclusion. An AI-generated score, ranking or narrative does not independently determine whether:

- A causal association exists
- A known risk has changed
- The information is clinically significant
- The benefit-risk balance has been affected
- Product information should be updated
- Additional pharmacovigilance is required
- Risk-minimisation measures are necessary
- Regulatory communication or escalation is appropriate

Precise terminology is a scientific and governance control. The expression “AI-detected signal” should be avoided when the system has generated only an alert, prioritised association or safety hypothesis.

3. Established Methodological Foundation

3.1 Disproportionality Screening

Established signal-detection programmes commonly combine quantitative screening with qualitative medical review. Methods used in spontaneous-reporting systems include the proportional reporting ratio, reporting odds ratio, information component and empirical Bayes approaches.[5,8,9]

Disproportionality measures use internal database comparator structures. They compare the proportion of reports containing a particular product-event combination with the corresponding reporting pattern in other parts of the database.

They do not:

- Estimate population incidence
- Provide a reliable exposed-person denominator
- Establish causality
- Operate as diagnostic tests
- Replace case-level medical review
- Resolve confounding or reporting bias

A signal of disproportionate reporting indicates that a product-event combination is reported more frequently than expected relative to the selected comparator. Its interpretation can be affected by publicity, regulatory action, product age, market penetration, indication, comorbidities, concomitant medication, coding conventions, duplicate reports, competition between events and changes in reporting behaviour.[5]

Thresholds therefore operate as screening criteria rather than decision rules. A positive disproportionality result can be clinically unimportant, while a medically important issue may arise from a small case series or a single well-documented report without meeting a quantitative threshold.

3.2 Broader Evidence Context

ICH E2E places pharmacovigilance planning in the context of the safety specification, identified risks, potential risks, missing information, epidemiology and limitations in the available human safety database.[6]

Signal assessment may require evidence from spontaneous reports, clinical trials, observational studies, mechanistic information, utilisation data, literature and regulatory experience. AI should be treated as an extension of this methodological environment, not as an independent replacement for it.

4. AI-Supported Functions

Artificial intelligence covers technologies with materially different behaviours, risks and validation requirements.

4.1 Rule-Based Automation

Rule-based applications execute predefined instructions, such as generating an alert when a threshold is exceeded or routing cases according to seriousness and expectedness.

They may improve consistency and efficiency, but they do not learn independently from data and should not automatically be described as AI. Their deterministic nature does not remove the need for software validation, change control and verification of rule logic.

4.2 Statistical Data Mining

Disproportionality calculations and related quantitative methods identify associations from reporting patterns. They remain central to signal detection but are not necessarily AI.

4.3 Supervised Machine Learning

Supervised models learn from labelled examples. They may classify or rank product-event combinations, cases or documents according to historical outcomes.

Their performance depends on the clinical validity, consistency and representativeness of the labels. Historical decisions may contain inconsistencies, process artefacts or information that was unavailable when an earlier decision was made.

4.4 Unsupervised Learning

Unsupervised methods identify clusters, anomalies or latent structures without predetermined outcome labels. They may reveal unusual case groupings or relationships that predefined queries do not capture.

The absence of labels does not reduce the need for validation. It increases the importance of clinical review, interpretability and controls against spurious clusters.

4.5 Natural Language Processing

Natural language processing may support:

- Identification of medicinal products and adverse events
- Extraction of temporality
- Recognition of dechallenge and rechallenge information
- Identification of risk factors and alternative explanations
- Classification of seriousness
- Literature screening
- Case clustering
- Structuring of narratives

Performance should be assessed at the level of the clinically relevant field or concept, not only at document level. Errors involving negation, temporality, experimenter, dose, product identity or causal language may have different clinical consequences.

4.6 Knowledge Graphs

Knowledge graphs can represent relationships among medicinal products, adverse events, indications, mechanisms, pathways and clinical concepts. They may support hypothesis generation and evidence navigation.

A represented relationship is not necessarily causal. The provenance, direction, confidence and date of each material relationship should remain accessible.

4.7 Generative AI

Generative AI functions should be separated because they have different hazards and acceptance criteria.

Retrieval

Output: Relevant controlled documents or passages.

Principal hazards: Missed evidence, irrelevant retrieval, outdated indexes, inappropriate access and malicious content.

Validation focus: Recall, precision, source currency, access controls, index completeness and malicious-document testing.

Extraction

Output: Defined facts copied or normalised from source records.

Principal hazards: Incorrect entity, negation, date, dose, patient or case attribution.

Validation focus: Field-level accuracy, source concordance, cross-case separation and clinically weighted error analysis.

Abstractive Summarisation

Output: A condensed narrative representing source evidence.

Principal hazards: Omission, distortion, unsupported synthesis, loss of uncertainty and failure to preserve contradictory evidence.

Validation focus: Factual consistency, omission testing, contradiction preservation, source-level citations and reviewer verification.

Recommendation Generation

Output: A suggested interpretation, priority or action.

Principal hazards: Hallucinated rationale, inappropriate authority, automation bias and insufficient abstention.

Validation focus: Clinical challenge testing, uncertainty communication, abstention performance, qualified review and explicit prohibition of autonomous signal or regulatory decisions.

Controls for generative AI should include:

- Exact source-level citations
- Verification against controlled source records
- Prevention and testing of cross-case contamination
- Prompt-injection and malicious-document testing
- Protection of patient and confidential information
- Versioning of the model, prompt, retrieval index and configuration
- Deterministic settings where feasible and justified
- Structured representation of uncertainty
- Abstention when evidence is insufficient
- Prevention and testing of invented references
- Validation of omissions and contradictions
- Retention of relevant input, output and review records
- Output that is repeatable or otherwise controlled to the degree necessary for the intended use

A fluent output is not evidence of factual reliability. Where a generated summary contributes to a safety decision, the reviewer must be able to trace every material statement to a controlled source.

5. Where AI Can Add Value

5.1 Data Preparation

AI-supported tools may assist with duplicate identification, terminology mapping, product normalisation, medical-concept extraction, detection of inconsistent information and conversion of unstructured text into structured variables.

These functions may improve the consistency of information presented to reviewers. They do not replace assessment of clinical relevance or source reliability.

5.2 Pattern Recognition

Machine-learning methods may examine multivariable patterns involving:

- Drug-drug-event relationships
- Comorbidities
- Subpopulations
- Time to onset
- Dose and duration
- Polypharmacy
- Repeated clinical trajectories
- Cross-source or cross-terminology relationships

These outputs remain hypotheses until medically and scientifically evaluated.

5.3 Alert Prioritisation

Prioritisation is one of the most plausible uses of AI. Relevant features may include seriousness, medical importance, fatal outcome, positive dechallenge or rechallenge, temporal plausibility, novelty, reporting trend, data completeness, vulnerable populations and cross-source consistency.

Prioritisation should direct expert attention. It should not silently exclude low-ranked information unless the exclusion has been separately justified, validated and controlled.

Safeguards for lower-ranked or suppressed outputs should include:

- Random sampling
- Risk-targeted sampling
- Specific review of rare and serious events
- Review of discordance between AI and conventional methods
- Predefined “never suppress” categories
- Follow-up of subsequently identified relevant issues
- Periodic assessment of threshold effects

5.4 Case-Series Analysis

AI may identify clinically similar cases and prepare structured case-series views covering patient characteristics, indication, dose, latency, treatment duration, concomitant medication, medical history, dechallenge, rechallenge, outcome and alternative aetiologies.

Every material summary statement should remain traceable to the source case. Reviewers must retain access to the complete underlying reports.

5.5 Literature Surveillance

AI may support title and abstract screening, evidence categorisation, extraction of study characteristics and retrieval of supporting or contradictory evidence.

A generated literature summary should not replace review of the source publication when its content contributes materially to signal validation, assessment or regulatory action.

5.6 Evidence Synthesis

A controlled evidence map may distinguish:

- Evidence supporting the hypothesis
- Evidence arguing against it
- Known confounders
- Alternative explanations
- Biological and pharmacological plausibility
- Epidemiological findings
- Missing information
- Unresolved contradictions

The system should differentiate retrieved evidence, extracted information, generated synthesis and human interpretation.

5.7 Workflow Orchestration

AI-supported systems may route alerts, assign tasks, monitor timelines, identify overdue activities, maintain version histories and support management information.

Workflow automation can provide substantial value without delegating scientific authority.

6. Scientific Evidence

6.1 Current Evidence Base

A 2025 critical review identified applications of machine learning and natural language processing across signal detection, validation and evaluation. Common signal-detection methods included clustering, random forests and gradient boosting.[10]

The scientifically appropriate formulation is:

Machine-learning approaches generally outperformed comparator disproportionality methods on study-specific retrospective metrics.[10]

This finding should not be extended beyond the available evidence. It does not establish:

- Prospective operational superiority
- Improved signal-management decisions
- Earlier regulatory action
- Improved benefit-risk decisions
- Better patient-safety outcomes

Methodological transparency in the reviewed literature was mixed, and only some studies used publicly available positive and negative control datasets.[10]

6.2 Evidence Ladder

Level 1: Technical Feasibility

Question: Can the technology perform the intended task under experimental conditions?

Evidence: Prototype testing, engineering verification and proof-of-concept studies.

Level 2: Analytical Validity

Question: Does the system produce sufficiently accurate and calibrated outputs on appropriate independent data?

Evidence: Locked testing, independent controls, temporal validation, calibration and subgroup analysis.

Level 3: Prospective Workflow Validity

Question: Does the system perform as intended with new data in the real or a realistically simulated workflow?

Evidence: Prospective silent-mode deployment and comparison with established surveillance.

Level 4: Operational Utility

Question: Does the combined workflow improve timeliness, consistency or resource use without creating unacceptable risk?

Evidence: Workload, reviewer time, alert quality, override behaviour, delays and process deviations.

Level 5: Patient-Safety or Regulatory Utility

Question: Does use contribute to better or earlier safety action or another defensible public-health benefit?

Evidence: Earlier medically relevant action, avoided delay, improved regulatory decisions or demonstrated patient-safety benefit.

Much of the published evidence remains concentrated in technical feasibility and retrospective analytical validity. Evidence of prospective workflow validity, operational utility and patient-safety or regulatory utility remains more limited.[10]

6.3 Evidence for Different Claims

Evidence that a model improves the ranking of known controls should include independent positive and negative controls, temporal separation, appropriate class-imbalance metrics, confidence intervals and product-held-out testing.

Evidence of earlier detection requires historically frozen data and labels, reconstruction of information availability over time, a predefined detection criterion and comparison of alert dates without future-data leakage.

Evidence that AI improves an operational pharmacovigilance process requires prospective workflow evaluation, workload and reviewer-time measures, human-factors assessment, missed-control follow-up, parallel surveillance and evaluation of decision quality and timeliness.

A model that ranks known controls effectively may still fail to detect emerging risks earlier. A model that generates earlier alerts may still fail to improve the operational process if it creates excessive workload, obscures uncertainty or adversely affects reviewer behaviour.

7. Failure Modes and Consequences

7.1 False-Positive Outputs

A highly sensitive model may identify large numbers of weak associations. Consequences include increased workload, prolonged review times, reduced attention to important information and alert fatigue.

The practical value of a system depends on the number and quality of outputs that reviewers can manage, not simply the proportion of known controls detected.

7.2 Missed or Delayed Issues

In open pharmacovigilance environments, the complete number of true negatives may not be known. The term “false-negative rate” should therefore be used only when the reference set, observation window, unit of analysis and criteria for a positive outcome are defined.

Where true negatives cannot be enumerated, more defensible measures include:

- Missed known controls
- Delayed detection
- Subsequently identified relevant issues
- Relevant cases incorrectly suppressed
- Serious or rare events omitted from review

Ownership for investigating missed or delayed issues must be assigned to a named function.

7.3 Bias and Representativeness

Pharmacovigilance data do not represent all patients equally. Differences in healthcare access, diagnosis, reporting behaviour, geography, language and data capture affect the available evidence.

Performance should be assessed across relevant products, therapeutic areas, age groups, sexes, geographical regions, languages, data sources, seriousness categories and event frequencies. Subgroup estimates should include confidence intervals where meaningful.

No dataset should be described as completely unbiased. The objective is to establish suitability, identify limitations and control the consequences of known or reasonably foreseeable bias.

7.4 Data Leakage and Overfitting

Performance can be inflated if future information, later label changes, duplicated cases or features derived from the outcome enter training or evaluation data. Historical labels may also reflect information unavailable when the alert would have needed to be generated.

Consequences include unreliable estimates of performance, poor generalisability and unjustified confidence in the system.

7.5 Model and Data Drift

Performance may change when reporting behaviour, products, terminology, data sources, clinical practice, user behaviour, model parameters or upstream systems change.

Drift can alter alert volume, calibration and subgroup performance without an immediately visible technical failure.

7.6 Limited Explainability

A model may be unsuitable even when its average discrimination is acceptable. If reviewers cannot understand the basis, limitations or uncertainty of an output, meaningful challenge becomes difficult.

The required degree of explanation depends on intended use and consequence. A tool used to organise work may require different controls from a tool used to suppress associations from further consideration.

7.7 Generative Errors and Malicious Inputs

Generative systems may invent statements or references, merge information from different cases, omit contradictory evidence, misunderstand negation or present uncertainty as fact. Retrieval-connected systems may also process malicious instructions embedded in documents.

Consequences include incorrect case attribution, privacy breaches, unsupported assessment language and contaminated decision records. Validation must therefore address omissions, contradictions, source attribution, cross-case separation and adversarial content.

7.8 Automation Bias

Reviewers may place excessive confidence in a numerical score or polished narrative. This risk increases when limitations are hidden, outputs appear authoritative or workload discourages independent review.

A high acceptance rate does not necessarily demonstrate good performance. It may also indicate ineffective challenge or excessive dependence on the system.

7.9 Accountability Gaps

Unclear boundaries among the marketing authorisation holder, technology provider, service provider and reviewer may delay incident investigation or safety decisions.

Contractual delegation does not remove the marketing authorisation holder's responsibility for its pharmacovigilance system.

7.10 System Unavailability

Outages, cybersecurity incidents, supplier failure, model withdrawal or loss of records may interrupt critical surveillance.

The organisation must be able to continue time-critical signal-management activities if the AI-supported system is unavailable, unreliable or suspended.

8. Regulatory and Scientific Context

8.1 Categories of Authority

AI governance should distinguish clearly among four categories.

Legal or regulatory requirements include applicable legislation and binding pharmacovigilance obligations.

Regulatory guidance and expectations include GVP modules, EMA questions and answers, reflection papers and FDA guidance. Their legal status depends on the instrument and jurisdiction.

Scientific good practice includes independent controls, temporal validation, calibration, prospective evaluation and clinically relevant subgroup testing.

Author recommendations include the HUMAN-SIGNAL Framework, the operating-autonomy model and the proposed inspection-ready evidence package.

8.2 EU Signal Management

GVP Module IX Revision 1 and Addendum I remain methodological foundations for signal management and signal detection from spontaneous reports. Later legislation and current EMA guidance affect the operational interpretation of sections concerning EudraVigilance monitoring and signal notification.[2-4]

EU Regulatory Update as of September 2026

Commission Implementing Regulation (EU) 2025/1466 amended Commission Implementing Regulation (EU) No 520/2012. The revised Article 18 requires marketing authorisation holders to monitor data available in EudraVigilance and use those data together with other available sources.[2]

Current EMA guidance states that all marketing authorisation holders with medicinal products authorised in the European Economic Area must use EudraVigilance as an additional source of safety information supporting their signal-management processes and enhancing signals detected through other sources.[3,4]

The former EudraVigilance monitoring pilot for marketing authorisation holders has ended. Following deletion of Article 21(2), marketing authorisation holders are no longer expected to submit validated signals through the former standalone signal-notification form as a general operating pathway.[2,3]

Signals identified through any source, including EudraVigilance, should be handled through the marketing authorisation holder's signal-management process. Applicable regulatory pathways remain available, including variation procedures, periodic safety update reports and notification of an emerging safety issue when the relevant criteria are met.[3,4]

Marketing authorisation holders should document how EudraVigilance data are used with other available sources, at which stages of the signal-management process they are considered and how the frequency and depth of monitoring reflect product characteristics, the known safety profile and risk.[3]

8.3 European Medicines Agency AI Position

The EMA reflection paper on AI in the medicinal-product lifecycle addresses development, performance assessment, interpretability, deployment, governance, data integrity and data protection. It includes post-authorisation activities such as adverse-event processing and signal detection.[11]

The joint EMA-FDA guiding principles of good AI practice emphasise human-centric design, risk-based controls, clear context of use, multidisciplinary expertise, data governance, appropriate performance assessment and lifecycle management. They state that risk-based performance assessment should evaluate the complete system, including human-AI interactions.[12]

These publications provide regulatory principles and expectations. They do not replace the identification of specific binding requirements applicable to a particular use case.

8.4 EU Artificial Intelligence Act

Regulation (EU) 2024/1689 establishes the European Union's horizontal legal framework for AI.[13] Regulation (EU) 2026/1744 amended parts of that framework, including the application timetable for requirements relating to high-risk AI systems.[14]

An AI system used in pharmacovigilance is **not automatically high-risk**. Classification depends on its intended purpose and whether it falls within an Article 6 pathway, including the product or safety-component pathway in Article 6(1) or a use case listed in Annex III under Article 6(2).

For each use case, the organisation should document:

- Whether the technology falls within the legal definition of an AI system
- Intended purpose and context of use
- System boundaries and excluded uses
- Provider, deployer, importer or other relevant operator roles
- Applicable Article 6 pathway
- Risk classification and rationale
- Applicable obligations and implementation dates
- Interactions with medicines, privacy, cybersecurity and employment law

The AI-literacy obligation requires providers and deployers to take measures, to their best extent, to ensure a sufficient level of AI literacy among relevant staff and other persons operating AI systems on their behalf.[13] Competence arrangements should therefore address intended use, limitations, source verification, uncertainty, automation bias, escalation and prohibited use.

Legal classification should be performed case by case with qualified legal and regulatory specialists. An internal pharmacovigilance risk rating does not replace classification under the AI Act.

8.5 United States Perspective

The FDA Center for Drug Evaluation and Research established the Emerging Drug Safety Technology Program to support discussion of research, development, validation and use of AI and other emerging technologies in post-market pharmacovigilance.[15]

Participation and discussions under the programme do not provide binding regulatory advice.[15] Sponsors and other eligible parties remain responsible for meeting applicable statutory and regulatory obligations.

FDA's January 2025 document, *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products*, remains draft and non-binding guidance. It proposes a context-of-use and risk-based credibility approach in which model influence and decision consequence inform the necessary level of credibility evidence.[16]

Although the document is not final guidance, its structure is relevant to pharmacovigilance AI governance. It distinguishes the question of interest, context of use, model risk, credibility plan, execution, documentation and determination of adequacy.

8.6 CIOMS Working Group XIV

The Council for International Organizations of Medical Sciences published the final report *Artificial Intelligence in Pharmacovigilance: Report of CIOMS Working Group XIV* in 2025.[17]

The report addresses AI across individual case safety report processing and signal management. It provides multidisciplinary consensus guidance rather than binding legal requirements.

9. Governance Controls

9.1 Complete Workflow Validation

The appropriate unit of validation is the complete socio-technical workflow, not the model in isolation.

The validated workflow should include:

- Data
- Model
- Software platform
- Interfaces and integrations
- Intended users
- Human-machine interactions
- Procedures
- Governance
- Documentation
- Resulting decisions

A technically accurate model may fail operationally because of poor interface design, misunderstood outputs, unavailable source evidence, inappropriate thresholds, workflow delays or uncritical human acceptance.

9.2 Decision Rights

AI may retrieve, extract, rank, route, draft or recommend.

Named and qualified humans must retain authority to:

- Validate signals
- Determine clinical relevance
- Approve signal-assessment conclusions
- Initiate or close escalation
- Determine benefit-risk implications
- Approve regulatory action

Human oversight must be substantive. Reviewers must have access to relevant source information, sufficient explanation of the output, suitable competence, authority to disagree and adequate review time.

9.3 The Author-Proposed HUMAN-SIGNAL Framework

HUMAN-SIGNAL is used as a governance mnemonic. Individual letters identify the relevant control domain but do not necessarily correspond to the first word of the full element title.

The HUMAN-SIGNAL Framework is an author-proposed practical governance structure. It is aligned with recognised pharmacovigilance and AI principles but is not a regulatory standard and does not replace applicable legal requirements.

HUMAN-SIGNAL Summary

Letter	Framework element	Key control question	Minimum evidence
H	Hypothesis, Not Conclusion	Is an AI output treated as a hypothesis until approved validation?	Alert status, validation decision and recorded rationale
U	Use Case and Intended Purpose	Are intended users, outputs, exclusions and autonomy defined?	Approved intended-use statement and fallback process
M	Measurable Performance	Are performance and operational error consequences measured?	Predefined measures, acceptance criteria and confidence intervals
A	Accountable Human Oversight	Who can review, override and make final decisions?	Decision-rights matrix, review and escalation records
N	Data Suitability, Representativeness and Bias Control	Are provenance, quality and foreseeable bias understood?	Data lineage, quality and representativeness assessment
S	Scientific Validation	Has the use case been tested on suitable independent and temporal evidence?	Validation protocols, controls, subgroup and silent-mode results
I	Inspection-Ready Documentation	Can system operation and resulting decisions be reconstructed?	Version, configuration, output, review and decision records
G	Governance Throughout the Lifecycle	Who owns change, quality, supplier and continuity controls?	Governance roles, supplier oversight and continuity plans
N	Ongoing Monitoring	Are drift, errors, overrides and missed controls followed?	Monitoring plan, control limits and performance reviews
A	Auditability and Explainability	Can an output be traced, explained and scientifically challenged?	Traceability, explanations and documented interpretation
L	Learning Under Controlled Change	Are material changes assessed, approved and reversible?	Impact assessment, regression test and rollback record

H: Hypothesis, Not Conclusion

AI outputs begin as algorithmic alerts, AI-prioritised associations or safety hypotheses. A statistical alert may support identification of a potential signal, but only the approved validation process can determine whether it becomes a validated signal. Signal assessment and regulatory conclusions remain distinct subsequent decisions.

U: Use Case and Intended Purpose

The approved intended-use statement should define:

- Problem and context of use
- Intended users
- Data sources
- Expected outputs
- Enabled actions
- Excluded and prohibited uses
- Decisions supported
- Decisions not permitted
- Autonomy level
- Human-review requirements
- Fallback process
- Foreseeable risks and expected benefits

The intended use should be locked before confirmatory validation. Material changes require impact assessment and, where necessary, revalidation.

M: Measurable Performance

Measures should reflect intended use, operational prevalence and the consequences of error. They may include:

- Sensitivity or recall
- Specificity where true negatives are definable
- Precision and positive predictive value
- Area under the precision-recall curve
- Area under the receiver operating characteristic curve where appropriate
- Calibration
- Confidence intervals
- Subgroup error
- Missed known controls
- Time to medically relevant alert
- Reviewer-time burden
- Alerts per reviewer-time unit
- Acceptance, modification and rejection rates

No single measure is sufficient.

A: Accountable Human Oversight

Governance should identify:

- Output reviewers
- Override authority
- Ownership of missed or delayed issues

- Disagreement-escalation routes
- Final decision rights
- Change approvers
- Incident owners
- Regulatory action authority

Responsibility should remain clear across the marketing authorisation holder, affiliates, service providers and technology suppliers.

N: Data Suitability, Representativeness and Bias Control

The mnemonic letter is retained, but the element does not imply that completely unbiased data can be achieved.

Assessment should cover:

- Provenance and lawful use
- Completeness and accuracy
- Coding consistency
- Duplicate handling
- Missingness
- Population representation
- Temporal relevance
- Geographic and linguistic coverage
- Label quality
- Known and foreseeable bias
- Changes over time

S: Scientific Validation

Validation should include, as appropriate:

- Locked intended use
- Predefined acceptance criteria
- Independent positive and negative controls
- Temporal validation
- External validation where feasible
- Product-held-out testing
- Therapeutic-area-held-out testing
- Frozen, historically appropriate labels and features
- Subgroup performance with confidence intervals
- Calibration assessment
- Precision-recall analysis
- Sensitivity analyses
- Failure-mode and challenge-case testing
- Prospective silent-mode evaluation
- Parallel manual surveillance during controlled rollout

Acceptance criteria should be approved before confirmatory testing.

I: Inspection-Ready Documentation

Documentation should support reconstruction of the system, its operation and the resulting decisions. Relevant records include intended use, requirements, data lineage, validation evidence, model or system cards, software and model versions, prompts, retrieval indexes, configurations, training records, output logs, decisions, overrides, deviations and performance reviews.

G: Governance Throughout the Lifecycle

Governance should cover development, implementation, operation, change and retirement. It should address ownership, provider and deployer roles, quality responsibilities, privacy, cybersecurity, supplier control, audit rights, records access, incident management and business continuity.

N: Ongoing Monitoring

The monitoring plan should define owners, cadence, indicators, control limits, corrective and preventive action triggers, suspension thresholds, escalation timelines and revalidation criteria.

Monitoring should address data drift, model drift, calibration, subgroup errors, alert volumes, override patterns, processing failures, user complaints and missed controls.

A: Auditability and Explainability

Four concepts should be distinguished:

- **Traceability:** The ability to reconstruct data, versions, actions and decisions.
- **Local explanation:** An explanation of a particular output.
- **Global model understanding:** An understanding of overall model behaviour and principal dependencies.
- **Scientific interpretability:** The ability to assess whether an output is clinically and pharmacologically meaningful.

The necessary degree of explanation depends on intended use, autonomy and consequence of error.

L: Learning Under Controlled Change

A production system should not change its behaviour without control. Change categories should include model retraining, prompt changes, threshold changes, terminology updates, retrieval-index changes, data-source additions, software updates, interface changes and supplier model substitutions.

Material changes should undergo impact assessment, regression testing, approval, controlled release and rollback planning.

9.4 Resilience and Fallback

Resilience and manual fallback apply across all eleven HUMAN-SIGNAL elements. The organisation must maintain the competence, data access and procedures needed to continue critical surveillance if the AI-supported system becomes unavailable, unreliable, compromised or suspended.

9.5 Alignment With Recognised Expectations

Hypothesis, Not Conclusion: Aligned with the GVP distinction between detection, validation, assessment and action.

Use Case and Intended Purpose: Aligned with FDA context-of-use concepts and the EMA-FDA principle of clear context of use.

Measurable Performance: Aligned with risk-based performance assessment, but expressed here as an author-proposed operational structure.

Accountable Human Oversight: Aligned with pharmacovigilance responsibilities and human-centric AI principles.

Data Suitability, Representativeness and Bias Control: Aligned with data-governance, provenance and fitness-for-use expectations.

Scientific Validation: Aligned with risk-based credibility and performance principles, but requiring use-case-specific implementation.

Inspection-Ready Documentation: A practical evidence structure supporting pharmacovigilance quality-system traceability.

Governance Throughout the Lifecycle: Aligned with lifecycle management, privacy, cybersecurity and supplier-control principles.

Ongoing Monitoring: Aligned with lifecycle monitoring and periodic re-evaluation.

Auditability and Explainability: A practical distinction among traceability, explanation, model understanding and scientific interpretation.

Learning Under Controlled Change: Aligned with change-control and lifecycle-management principles.

This mapping indicates alignment, not equivalence. HUMAN-SIGNAL is not an official regulatory framework.

10. Assist-Augment-Orchestrate Operating-Autonomy Model

The Assist-Augment-Orchestrate Operating-Autonomy Model describes increasing operational involvement, not increasing organisational maturity.

10.1 Assist

The system supports relatively low-autonomy tasks such as:

- Literature triage
- Duplicate suggestions
- Case clustering
- Defined data extraction
- Retrieval of related evidence
- Draft summaries

Human review remains extensive. The system does not determine whether an alert proceeds within signal management.

10.2 Augment

The system performs more active analytical support through:

- Alert prioritisation
- Multi-source pattern identification
- Suggested evidence maps
- Identification of supporting and contradictory evidence
- Structured case-series analysis
- Draft signal-assessment components

Qualified pharmacovigilance professionals continue to validate, prioritise, assess and approve decisions.

10.3 Orchestrate

The system coordinates information and workflow across multiple activities. It may:

- Integrate outputs from multiple sources
- Route potential issues
- Monitor status and deadlines
- Identify dependencies
- Maintain structured decision histories
- Trigger pre-approved administrative steps

Scientific conclusions and regulatory decisions remain under accountable human control.

Greater autonomy is not equivalent to greater maturity.

A mature organisation may deliberately retain low autonomy for a high-consequence use case because uncertainty, limited explainability or the potential impact of an incorrect output requires intensive human control.

11. Metrics and Acceptance Criteria

11.1 Statistical Measures

Accuracy may be misleading where relevant positive outcomes are rare. Performance measures should be selected according to the task, reference set and operational consequences.

Sensitivity or recall measures the proportion of defined positive reference items detected.

Specificity is appropriate only where a meaningful true-negative set can be enumerated.

Precision or positive predictive value measures the proportion of alerts that meet the defined positive reference.

Area under the precision-recall curve is particularly relevant for strongly imbalanced classification tasks.[18]

Area under the receiver operating characteristic curve may be useful for discrimination but can appear favourable in highly imbalanced datasets.

Calibration evaluates agreement between predicted probabilities and observed outcomes. It should be assessed separately from discrimination.[19]

Confidence intervals indicate the uncertainty associated with performance estimates.

Subgroup error identifies performance variation across clinically relevant populations, products or data sources.

Time to medically relevant alert measures timeliness from the availability of qualifying evidence to the generation of an actionable alert.

11.2 Operational Measures

Operational evaluation should include:

- Alerts per reviewer-time unit
- Review time per alert
- Time from alert to documented decision
- Total workload
- Duplicate work

- Backlog and overdue activities
- Outputs accepted, modified or rejected
- Reviewer agreement
- Escalation timeliness
- System availability
- Manual-fallback performance

11.3 Human-Factors Measures

Human-factors assessment should include:

- Override frequency and rationale
- Uncritical acceptance patterns
- User confidence calibration
- Alert fatigue
- Differences among reviewers
- Source-verification behaviour
- Training and competence
- Retention of medical and scientific review skills

A low override rate is not necessarily evidence of good system performance. It may indicate an accurate system, ineffective oversight or automation bias.

11.4 Acceptance Criteria

Acceptance criteria should be predefined and linked to the intended use. They should address:

- Minimum performance on independent controls
- Maximum acceptable missed-control rate
- Operational alert burden
- Required calibration
- Subgroup performance limits
- Maximum acceptable delay
- Source-traceability requirements
- Critical extraction-error limits
- Availability and recovery targets
- Conditions requiring corrective action, suspension or rollback

A system should not pass validation solely because an average metric exceeds a threshold. Material failure in a clinically important subgroup or “never suppress” category may be disqualifying.

12. Implementation Sequence

12.1 Define the Use Case

Start with a specific operational problem rather than a general objective to implement AI. Define context of use, users, data, outputs, autonomy, prohibited uses, expected benefit and consequences of error.

12.2 Establish Governance

Assign process, scientific, technical, quality, privacy, security and supplier responsibilities. Determine relevant provider and deployer roles and obtain case-specific legal analysis.

12.3 Assess the Data

Document provenance, permissions, lineage, quality, missingness, coding, duplicates, representativeness, temporal relevance and foreseeable bias.

12.4 Design the Controls

Specify human review, source access, explanations, thresholds, “never suppress” categories, lower-ranked sampling, incident management, suspension criteria and manual fallback.

12.5 Validate Analytically

Use locked requirements, independent controls, predefined criteria, temporally appropriate data and clinically relevant subgroup and calibration analyses.

Sensitivity analyses should address:

- Duplicate handling
- Terminology versions
- Notoriety bias
- Stimulated reporting
- Product age
- Changes in reporting
- Alternative thresholds
- Missing-data handling
- Label uncertainty

12.6 Test Prospectively in Silent Mode

Operate the system prospectively without allowing it to control production decisions. Compare its outputs with the established surveillance process using new data.

Silent-mode evaluation should assess:

- Missed or delayed known controls
- Alert burden
- Ranking stability
- Clinically relevant discordance
- Subgroup errors
- Reviewer-time impact
- Data and integration failures

12.7 Use a Controlled Rollout

During initial deployment, maintain parallel manual surveillance proportionate to the risk. Define rollout duration, enhanced monitoring, escalation criteria, suspension thresholds, rollback triggers and the evidence required before reducing parallel controls.

12.8 Monitor and Control Change

Monitor technical, scientific, operational and human-factors performance. Changes should proceed through classification, impact assessment, testing, approval, release and rollback.

12.9 Retire Responsibly

Retirement planning should preserve records, data access, decision histories and regulatory reconstruction capability. Operational continuity should be established before discontinuing a system or supplier.

13. Inspection-Ready Implementation

An inspection-ready implementation should demonstrate that the organisation understands the use case, system boundaries, data, risks, human responsibilities and evidence supporting continued operation.

The core evidence should cover:

- Intended and prohibited uses
- System boundaries and data flows
- Data suitability and lineage
- Risk and legal-role assessments
- Requirements and validation
- Human factors and decision rights
- Supplier oversight
- Audit trails
- Performance monitoring
- Change control
- Business continuity
- Retirement

The complete evidence package is listed in Appendix A.

For each material AI-supported output, the organisation should be able to reconstruct the source information available at the time, relevant transformations, model and software versions, configuration, output presented to the reviewer, human review, overrides, final rationale, escalation and resulting action.

Supplier contracts should support appropriate oversight. Relevant provisions should address changes, subcontractors, security, incident notification, data retention, audit and inspection support, access to records, transition and exit assistance.

14. Practical Recommendations for Marketing Authorisation Holders

14.1 Start With the Problem

Define the process problem, intended context, expected benefit and consequence of an incorrect output before selecting a technology.

14.2 Preserve Regulatory Terminology

Procedures, systems and training should distinguish algorithmic alerts, AI-prioritised associations, safety hypotheses, potential signals, validated signals, assessment conclusions and regulatory decisions.

14.3 Validate the Workflow

Do not validate only the model. Include data ingestion, terminology mapping, interfaces, user interaction, output presentation, human review, documentation, escalation and resulting decisions.

14.4 Use Prospective Evidence

Retrospective testing is necessary but not sufficient. Prospective silent-mode testing should establish how the system performs with new data under realistic operational conditions.

14.5 Test Failure Modes Deliberately

Testing should include incomplete cases, conflicting dates, duplicates, rare events, multilingual content, unusual terminology, confounding, misclassified products, vulnerable populations, stimulated reporting and changing data distributions.

14.6 Control Generative AI Separately

Retrieval, extraction, summarisation and recommendation generation should have separate requirements, hazards, validation methods and acceptance criteria.

14.7 Protect Lower-Ranked Information

Use random sampling, risk-targeted sampling, rare and serious event review, discordance analysis and “never suppress” categories to identify weaknesses in prioritisation or suppression.

14.8 Monitor Human Behaviour

Assess whether reviewers become overly dependent on recommendations, disregard low-ranked information, reduce source verification or lose important medical-review skills.

14.9 Maintain Manual Capability

Business continuity should include the ability to continue critical signal-management activities if the AI-supported system is unavailable, unreliable or suspended.

14.10 Require Supplier Transparency

The marketing authorisation holder should understand relevant data handling, model versions, changes, subcontracting, security controls, performance monitoring, incident notification, retention, discontinuation and access-to-records arrangements.

14.11 Treat Validation as a Lifecycle

Initial validation is the starting point. Continued confidence requires monitoring, periodic review, controlled change, appropriate revalidation and defined suspension or retirement criteria.

15. Discussion

AI can add value where pharmacovigilance involves large volumes, repetitive classification, complex comparison, evidence retrieval or workflow coordination. Its most credible role is often to improve how information is prepared and prioritised rather than to automate final scientific decisions.

The available literature indicates technical promise. Machine-learning approaches have performed well against traditional comparators on study-specific retrospective metrics, and natural language processing has supported extraction and validation-related tasks. These findings justify further controlled implementation, not assumptions of general superiority.[10]

An algorithm may be sensitive but operationally unsuitable because it generates too many weak alerts. It may perform well on average but fail for rare events or underrepresented groups. It may reproduce

historical decisions while also reproducing historical inconsistencies. A generative model may produce a persuasive summary while omitting contradictory evidence.[20]

Human oversight cannot function as ceremonial approval. Reviewers must be able to understand the role and limitations of the output, examine the relevant source evidence, disagree with the recommendation and document the final scientific rationale.

Machines are well suited to systematic retrieval, large-scale comparison, repetitive classification and computational pattern recognition. Human experts remain essential for clinical context, causal reasoning, interpretation of uncertainty, evaluation of alternative explanations and accountable regulatory decision-making.

The objective is not to allocate every possible task to AI. It is to allocate tasks in a manner that improves the reliability, timeliness and transparency of the complete signal-management process.

16. Conclusion

Artificial intelligence will not remove uncertainty from pharmacovigilance. It cannot transform incomplete observational information into definitive causal evidence, and it cannot eliminate the need for medical and scientific judgement.

AI can make complex information more accessible, identify patterns that merit attention, improve prioritisation, support review of large case series and organise evidence across multiple sources. Used within defined boundaries, it may contribute to earlier, more consistent and more transparent safety assessment.

These benefits cannot be assumed from retrospective model performance. They require locked intended use, suitable data, predefined acceptance criteria, independent and temporal validation, prospective workflow evaluation, meaningful human oversight, controlled change and continuous monitoring.

The future of signal detection is neither fully manual nor presumptively autonomous. It is a human-governed system in which validated technologies undertake suitable computational tasks while qualified pharmacovigilance professionals remain accountable for clinical interpretation, signal validation, signal assessment, benefit-risk conclusions and regulatory action.

Success must be judged by the quality, timeliness, traceability and safety of the complete human-machine process.

AI should not replace the scientific process of signal management.

It should make that process better.

Appendix A: Inspection-Ready Evidence Package

A proportionate evidence package should include:

1. Intended-use statement
2. Prohibited-use statement
3. Context-of-use description
4. System-boundary and data-flow diagram
5. Algorithmic impact and risk assessment
6. AI Act and other applicable legal-role assessment
7. Data inventory and lineage
8. Data-quality and representativeness assessment
9. Privacy and information-security assessment
10. User, functional and performance requirements

11. Validation plan, protocols, datasets and report
12. Human-factors and automation-bias assessment
13. Model or system card
14. Prompt, retrieval-index and configuration records where applicable
15. Supplier qualification and supporting evidence
16. Contractual audit, inspection and records-access rights
17. Responsibility and decision-rights matrix
18. User instructions, training and competence records
19. End-to-end audit trail
20. Performance-monitoring plan
21. Incident, deviation and corrective and preventive action procedures
22. Change-control, release, suspension and rollback records
23. Business-continuity, manual-fallback, periodic-review and retirement plans

References

1. European Medicines Agency; Heads of Medicines Agencies. [Guideline on good pharmacovigilance practices: Module IX, Signal management](#). EMA/827661/2011 Rev. 1. Final version adopted 9 October 2017; effective 22 November 2017. Accessed 29 September 2026.
2. European Commission. [Commission Implementing Regulation \(EU\) 2025/1466 of 22 July 2025 amending Implementing Regulation \(EU\) No 520/2012 as regards pharmacovigilance activities](#). Off J Eur Union. 2025;L 2025/1466. Accessed 29 September 2026.
3. European Medicines Agency. [Questions and Answers on Implementing Regulation \(EU\) 2025/1466: Amendment of Regulation \(EU\) No 520/2012 and Conclusion of the Signal Detection in EudraVigilance Pilot by MAHs](#). EMA/243145/2025 Rev. 2. 28 January 2026. Accessed 29 September 2026.
4. European Medicines Agency. [Questions and answers on signal management](#). EMA/261758/2013 Rev. 5. 19 January 2026. Accessed 29 September 2026.
5. European Medicines Agency; Heads of Medicines Agencies. [Guideline on good pharmacovigilance practices: Module IX Addendum I, Methodological aspects of signal detection from spontaneous reports of suspected adverse reactions](#). EMA/209012/2015. Final version adopted 9 October 2017; effective 22 November 2017. Accessed 29 September 2026.
6. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. [ICH Harmonised Tripartite Guideline E2E: Pharmacovigilance Planning](#). Step 4 version. 18 November 2004. Accessed 29 September 2026.
7. Uppsala Monitoring Centre. [About VigiBase](#). Accessed 29 September 2026.
8. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483-486. [doi:10.1002/pds.677](#).
9. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998;54(4):315-321. [doi:10.1007/s002280050466](#).
10. Warner J, Prada Jardim A, Albera C. Artificial intelligence: applications in pharmacovigilance signal management. *Pharmaceut Med*. 2025;39(3):183-198. [doi:10.1007/s40290-025-00561-2](#).
11. European Medicines Agency. [Reflection paper on the use of artificial intelligence in the medicinal product lifecycle](#). EMA/CHMP/CVMP/83833/2023. Final version adopted by the Committee for Medicinal Products for Human Use on 9 September 2024 and by the Committee for Veterinary Medicinal Products on 11 September 2024. Accessed 29 September 2026.
12. European Medicines Agency; US Food and Drug Administration. [Guiding principles of good AI practice in drug development](#). January 2026. Non-binding guiding principles. Accessed 29 September 2026.

13. European Parliament; Council of the European Union. [Regulation \(EU\) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence](#). Off J Eur Union. 2024;L 2024/1689. Accessed 29 September 2026.
14. European Parliament; Council of the European Union. [Regulation \(EU\) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations \(EU\) 2024/1689, \(EU\) 2018/1139 and \(EU\) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence](#). Off J Eur Union. 2026;L 2026/1744. Accessed 29 September 2026.
15. US Food and Drug Administration. [Emerging Drug Safety Technology Meetings: Program Announcement](#). Fed Regist. 2024;89:49179-49180. Document 2024-12770. Accessed 29 September 2026.
16. US Food and Drug Administration. [Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products](#). Guidance for Industry and Other Interested Parties. Draft guidance, not for implementation. January 2025. FDA document 59502407dft. Accessed 29 September 2026.
17. Council for International Organizations of Medical Sciences. Artificial intelligence in pharmacovigilance: report of CIOMS Working Group XIV. Geneva: CIOMS; 2025. ISBN 978-92-9036-110-7. doi:10.56759/cdob6397.
18. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the receiver operating characteristic plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432. doi:10.1371/journal.pone.0118432.
19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. doi:10.1186/s12916-019-1466-7.
20. National Institute of Standards and Technology. [Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile](#). NIST AI 600-1. Gaithersburg, MD: National Institute of Standards and Technology; 2024. doi:10.6028/NIST.AI.600-1.