ClinSphere Journals
International Peer-Reviewed Open Access Journal
Advancing Healthcare, Clinical Research & Innovation
DOI: 10.5281/zenodo.23794745

Review Article

Artificial Intelligence in Pharmacovigilance: Automated Adverse Drug Reaction Detection and Signal Management

Authors: Dr. Vrushti Mitesh Shah

Affiliation: Clinical Pharmacist, Dhiraj Hospital, Vadodara, Gujarat, India

**Corresponding author: Dr. Vrushti Shah (dr.vrushtishah25@gmail.com)*

Acknowledgments: None.

ABSTRACT

Artificial intelligence (AI), including natural language processing, machine learning, and generative models, can assist pharmacovigilance by finding and structuring safety information across individual case safety reports (ICSRs), clinical records, scientific literature, and other sources. This narrative review examines how such tools support adverse drug reaction (ADR) detection and signal management, the evidence for operational use, and the safeguards required for responsible deployment. Current applications include report intake and duplicate detection, extraction of drugs and events from narrative text, case triage, literature screening, coding support, and prioritization of drug-event pairs for expert assessment. Published studies and a systematic review demonstrate technical feasibility, but evidence is heterogeneous and often task-specific; accuracy on curated datasets does not establish generalizability, causal association, or improved patient outcomes. Spontaneous-report disproportionality remains hypothesis-generating and is vulnerable to underreporting, stimulated reporting, missing denominators, and confounding. AI outputs therefore require documented context of use, representative validation, human review proportionate to risk, privacy protection, bias surveillance, change control, and traceable records. Current expectations include EMA's 2024 lifecycle reflection paper, the 2025 ICH E2D(R1) guideline effective from 18 March 2026, and FDA–EMA Good AI Practice principles issued in 2026. AI is best positioned as governed decision support within the established signal-management process, with pharmacovigilance professionals retaining responsibility for interpretation and action.

Keywords: Ob pharmacovigilance; artificial intelligence; machine learning; natural language processing; adverse drug reaction; signal detection; safety surveillance

1. Introduction

Pharmacovigilance (PV) is the science and set of activities concerned with detecting, assessing, understanding, and preventing adverse effects and other medicine-related problems. Post-authorisation surveillance must integrate incomplete reports, heterogeneous clinical narratives, evolving medical terminology, and evidence that arrives at different times. Human review remains essential, yet the volume and diversity of safety data create opportunities for carefully scoped automation. [1,2]

AI is an umbrella term rather than a single method. In PV, it includes rule-based systems, statistical learning, supervised and unsupervised machine learning, deep neural networks, NLP, and generative AI/large language models (LLMs). The practical question is not whether a model is "intelligent," but whether a defined task—such as screening literature or suggesting a MedDRA term—can be performed reliably in a specified workflow, with errors detected and accountable review preserved. [1,3]

This review focuses on AI-assisted ADR detection and safety signal management in post-marketing settings. It separates operational automation (finding, extracting, classifying, and prioritising evidence) from signal evaluation and regulatory or clinical decisions. A statistical association or model alert is not, by itself, an ADR diagnosis, a causal conclusion, or a validated signal.

2. Methodology / Literature Search Strategy

This article is a narrative review, not a systematic review. A targeted search was conducted on 30 October 2026 using PubMed-indexed and scholarly search results, official EMA, FDA, ICH, and CIOMS materials, and reference lists of relevant reviews. Search concepts combined "pharmacovigilance" or "drug safety" with "artificial intelligence," "machine learning," "natural language processing," "adverse drug reaction," "signal detection," and "large language model." Priority was given to current authoritative guidance, systematic reviews, and primary studies that illustrate established technical tasks or implementation considerations.

No reproducible database-by-database search string, dual screening, formal risk-of-bias assessment, protocol registration, or quantitative synthesis was performed. The selection is therefore selective and may omit relevant studies; the article does not claim exhaustive coverage or systematic-review certainty. Regulatory documents are described according to their status and publication date, and research performance figures are presented only with their task and dataset context.

3. Background

Signal management is a continuous, documented sequence: detection, validation, prioritisation, assessment, recommendation for action, and communication. In the EU,

EMA Good Pharmacovigilance Practices (GVP) Module IX describes this lifecycle, while the statistical addendum discusses quantitative methods. Screening tools can help identify patterns, but the signal must be validated and assessed using all relevant evidence, including case narratives, exposure, biologic plausibility, clinical context, and alternative explanations. [4]

AI can contribute at three connected layers. First, it can process source material: identify potential cases, detect duplicates, classify seriousness or reportability, and extract drug-event-time relationships. Second, it can structure evidence: standardize terminology, map products and reactions, summarize narratives, and link reports to literature or clinical records. Third, it can rank candidate drug-event pairs using statistical, temporal, semantic, or predictive features. The output at each layer is a work aid whose intended use and failure costs differ.

Spontaneous reporting systems (SRS) such as FAERS, EudraVigilance, and VigiBase are central to signal generation but do not provide complete population exposure denominators. Reporting odds ratios, proportional reporting ratios, and Bayesian shrinkage statistics quantify disproportionate reporting under particular assumptions; they do not estimate incidence or prove causality. Results depend on data cleaning, duplicate handling, coding choices, comparator selection, and reporting dynamics. [4,5]

4. Current Evidence

4.1 Case intake, processing, and quality control

NLP and classification models can identify likely safety reports in free text, extract suspect products and adverse events, classify report types, flag missing elements, and suggest coding or routing. In a literature-monitoring case study, a validated classifier filtered 55% of irrelevant articles while targeting 99% recall for suspected adverse-event articles; this is an operational example, not evidence that all sources or PV systems will achieve the same performance. [6] Automated extraction can reduce repetitive review, but source documents and original language must remain available for verification.

4.2 ADR detection in unstructured text

Social media, online drug reviews, clinical notes, and published literature contain patient language that differs from standardized case terminology. NLP pipelines commonly combine named-entity recognition, relation extraction, negation and temporality detection, normalization to drug and reaction vocabularies, and document-level classification. Review evidence indicates diverse use cases, with much of the published AI literature focused on identifying adverse drug events/reactions and processing reports; studies vary substantially in data source, annotation, and endpoint. [7,8]

Technical benchmarks often use precision, recall, F1 score, or area under the receiver operating characteristic curve. For PV, recall and workload reduction may matter together: a screen that removes irrelevant documents but misses rare serious reports can be unsafe. Performance should be assessed at the relevant operational threshold, with false negatives, subgroup results, calibration, and downstream reviewer workload reported. External and temporal validation are particularly important when language, products, data capture, or coding practice changes.

4.3 Signal detection and prioritisation

AI methods can supplement established disproportionality methods by incorporating report-level features, temporal trends, medical concepts, co-reported medicines, or cross-source evidence. Machine learning can identify nonlinear patterns and rank alerts for review, while NLP can expand evidence beyond coded event tables. However, model ranking may inherit reporting biases and label limitations. An "important" feature is predictive within a dataset, not proof of pharmacologic mechanism or causality.

Signal detection remains an evolving statistical discipline. Established quantitative methods are useful for screening spontaneous reports, but their outputs require clinical and epidemiologic assessment. [5] AI should therefore be benchmarked against fit-for-purpose baselines—such as existing rules, disproportionality metrics, and expert-reviewed workflows—and evaluated for incremental value, not merely standalone accuracy. A candidate signal should be replicated or investigated in independent data where possible and assessed with appropriate pharmacoepidemiologic designs. [5,12]

4.4 Generative AI and large language models

Generative models may assist with narrative summarisation, structured data extraction, literature triage, query generation, and drafting preliminary case narratives. Their flexibility is accompanied by risks of fabricated details, omission, altered certainty, inconsistent terminology, and privacy leakage. A proof-of-concept guardrail approach for translating structured and unstructured ICSR content illustrates possible controls, including anomaly detection, terminology validation, and explicit uncertainty; it does not establish autonomous case processing as safe. [9] A human reviewer must verify every safety-relevant generated statement against source evidence.

5. Clinical and Operational Applications

PV workflow	AI-supported task	Potential value	Required control
Safety intake	Detect potential ICSRs in email, web forms, call transcripts, literature	Earlier routing; less manual screening	High-recall validation; missed-case monitoring; source traceability
Case processing	Extract drug, event, dates, seriousness, outcome; detect duplicates	Faster structured entry and QC	Field-level evidence links; reviewer confirmation; audit trail
Coding and triage	Suggest MedDRA terms; classify seriousness or expectedness for review	Consistent prioritisation	Version control; terminology governance; no unchecked coding

Literature monitoring	Rank abstracts/full text for potential safety evidence	Reduce irrelevant screening volume	Recall target set by risk; monitor drift and missed relevant articles
Signal screening	Combine disproportionality, text, temporal and clinical features	Prioritise candidate pairs	Compare with baseline; independent validation; alert adjudication
Aggregate review	Retrieve and summarise evidence for periodic safety reports	Support synthesis and traceability	Citations to source records; human authorship and sign-off
Post-market follow-up	Identify cohorts or subgroups for targeted assessment	Generate focused hypotheses	Privacy, bias assessment, study design and confounding control

Table 1. Illustrative AI-enabled tasks across the PV workflow. "Potential value" describes plausible operational contribution; it does not imply proven improvement in patient outcomes.

6. Safety and Limitations

Data quality and representativeness. Incomplete, duplicated, delayed, and selectively reported data can distort both model inputs and labels. A model trained on one reporting system, language, geography, or product class may not transfer. Clinical notes and social media can introduce additional selection and verification problems. Data provenance, permitted use, data lineage, and source-level uncertainty should be documented.

Bias and unequal error burden. Underrepresentation of age groups, pregnancy, rare diseases, languages, or low-resource settings can reduce sensitivity for those populations. Aggregate performance may conceal subgroup failures. Teams should test relevant strata, investigate differential false negatives and false positives, and establish escalation routes when evidence is sparse.

Causal overinterpretation. Neither a disproportionate reporting statistic nor a high model score demonstrates that a medicine caused an event. Confounding by indication, notoriety, concomitant therapy, stimulated reporting, exposure changes, and duplicate follow-up reports can generate apparent patterns. Signal evaluation must use clinical review and suitable epidemiologic evidence, with explicit uncertainty.

Automation bias and workflow risk. Reviewers may over-trust a confident score or miss details that the model did not surface. Conversely, frequent low-value alerts can create fatigue. Interfaces should display evidence, uncertainty, provenance, and a practical route to inspect the source. Human oversight must be calibrated to the use case: low-risk document routing differs from serious-event assessment or a recommendation that could change labeling.

Model drift, explainability, and change control. Terminology, product portfolios, source channels, and reporting practices evolve. Monitoring should include data drift and outcome-based performance, with predefined thresholds, retraining governance, version history, rollback, and change impact assessment. Interpretability tools can assist review but do not guarantee that a model is clinically valid.

Privacy, security, and vendor risk. Sensitive health data require lawful, secure handling and access control. External generative services may retain inputs or expose confidential information; use requires privacy review, contractual safeguards, and technical controls. PV quality systems should account for model components, data pipelines, validation evidence, user training, deviations, and auditability.

7. Emerging Advances

Multimodal models may connect structured safety tables with narratives, laboratory results, imaging reports, and published evidence. Retrieval-augmented generation can ground summaries in approved source documents and return citations for verification. Federated learning and privacy-preserving methods could enable cross-institutional model development without centralizing raw data, although governance and harmonized labels remain challenging. Knowledge graphs may help represent drug-event-indication relationships and support traceable evidence navigation.

These advances should be treated as research directions. Robust prospective studies are needed to determine whether new architectures improve sensitivity, review time, consistency, or decision quality over established tools. Benchmark datasets should include temporal and geographic holdouts, realistic class imbalance, source provenance, adjudication procedures, and transparent error analysis.

8. Pharmacist's Role

Pharmacists bring expertise in medication histories, pharmacology, interactions, therapeutic context, and clinically meaningful event interpretation. In AI-assisted PV, their contribution spans design, review, governance, and communication. They can define clinically relevant extraction fields; review drug-event temporality, dose, dechallenge/rechallenge, concomitant medicines, and alternative causes; challenge implausible model outputs; and identify patterns that warrant further assessment.

therapy.

Responsibility	Practical pharmacist contribution
Use-case design	Define the PV task, intended users, error costs, escalation rules, and acceptance criteria.
Clinical verification	Check the source narrative, medication chronology, event coding, seriousness, and clinical plausibility.
Signal assessment	Interpret candidate associations alongside exposure, indication, mechanism, confounding, and external evidence.
Quality and governance	Review validation plans, subgroup performance, drift reports, SOPs, audit trails, and change controls.
Team and patient communication	Explain uncertainty, support staff training, and ensure decisions and safety communications are understandable.

Table 2. Pharmacist contributions to safe and accountable AI-supported pharmacovigilance.

9. Research Gaps

The literature remains uneven across tasks, data sources, and outcomes. Many studies are retrospective and use curated or single-source datasets; external validation and prospective workflow evaluation are less common. Reporting of annotation methods, missing data, class imbalance, operating thresholds, calibration, subgroup performance, and model updates is inconsistent. Published accuracy metrics often do not measure the operational harm of missed serious cases or the consequences of excess alerts. [7,13]

Further gaps include harmonized benchmarks for multilingual and cross-jurisdictional PV, transparent handling of duplicate and follow-up reports, methods for causal evidence integration, evaluation of human-AI interaction, privacy-preserving data linkage, and independent reproducibility. Most importantly, the field needs evidence that a deployed system improves timeliness, completeness, signal quality, workload, or patient-relevant outcomes without weakening detection or accountability.

10. Future Perspectives

The most defensible near-term strategy is staged deployment. Begin with a narrowly defined task where source evidence is accessible and a human can readily verify the output. Establish a baseline workflow, prespecify acceptable errors and performance thresholds, validate on representative data including temporal and external holdouts, and prospectively monitor real-world performance. Expand scope only after the system demonstrates measurable benefit and the safety case is maintained.

Governance should follow risk and context of use. EMA's 2024 reflection paper addresses AI across the medicinal-product lifecycle; FDA's 2025 draft guidance proposes a risk-based credibility assessment for AI supporting regulatory decision-making; and FDA–EMA Good AI Practice principles published in January 2026 emphasize human-centric design, clear context of use, risk management, standards, and lifecycle oversight. [2,3,10] ICH E2D(R1), effective 18 March 2026, updates post-approval safety data definitions and ICSR management/reporting standards, including contemporary sources. These frameworks complement rather than replace PV-specific obligations and quality-system controls. [11]

Future systems should make outputs auditable: retain the input and source location, model and data versions, confidence or uncertainty, user changes, and final decision rationale. Signal management remains a multidisciplinary responsibility. AI can increase the reach and speed of screening, but scientific assessment, regulatory judgment, and accountability must remain with qualified professionals.

11. Conclusion

AI offers practical support for detecting potential ADR information, structuring reports, screening literature, and prioritising safety hypotheses. The evidence supports task-specific technical feasibility, not a general claim that AI independently discovers causal ADRs or improves patient outcomes. Safe value depends on rigorous validation, transparent limits, representative data, privacy and bias controls, lifecycle monitoring, and expert review embedded in existing signal-management systems. Pharmacists and other PV professionals remain central to converting model outputs into defensible safety assessments and proportionate action.

12. Declarations

Funding: No external funding was provided for preparation of this review.

Conflicts of interest: The author(s) declare no competing interests. This statement must be confirmed and amended by the submitting author(s) before submission.

Ethics approval: Not applicable; this narrative review used published literature and public guidance and involved no human participants or identifiable private data.

Data availability: No new datasets were generated or analyzed for this review.

Author contributions: Author contributions were not supplied. Complete this declaration with the actual contributors and their roles before submission; no authors are invented here.

AI-use declaration: Generative AI assistance was used to support literature discovery, drafting, and document preparation. The submitting author(s) must independently verify all claims, references, and declarations and disclose AI use according to the target journal's current policy.

13. References

1. Council for International Organizations of Medical Sciences (CIOMS). Artificial intelligence in pharmacovigilance: report of the CIOMS Working Group XIV. Geneva: CIOMS; 2025.

2. European Medicines Agency. Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. EMA/CHMP/CVMP/83833/2023. Amsterdam: EMA; 2024.

3. US Food and Drug Administration, European Medicines Agency. Guiding principles of good AI practice in drug development. Silver Spring (MD): FDA; Amsterdam: EMA; 2026 Jan.

4. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): Module IX—Signal management (Rev 1). EMA/827661/2011 Rev 1. London: EMA; 2017.

5. Bate A, Evans SJW. Quantitative signal detection using spontaneous adverse drug reaction reporting. Pharmacoepidemiol Drug Saf. 2009;18(6):427-436. doi:10.1002/pds.1742.

6. Ohana B, Sullivan J, Baker N. Validation and transparency in AI systems for pharmacovigilance: a case study applied to the medical literature monitoring of adverse events. arXiv [Preprint]. 2022. arXiv:2201.00692. doi:10.48550/arXiv.2201.00692.

7. Salas M, Petracek J, Yalamanchili P, Aimer O, Kasthuril D, Dhingra S, et al. The use of artificial intelligence in pharmacovigilance: a systematic review of the literature. Pharm Med. 2022;36:295-306. doi:10.1007/s40290-022-00441-z.

8. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. Drug Saf. 2016;39(3):231-240. doi:10.1007/s40264-015-0379-0.

9. Hakim JB, Painter JL, Ramcharran D, Kara V, Powell G, Sobczak P, et al. The need for guardrails with large language models in medical safety-critical settings: an artificial intelligence application in the pharmacovigilance ecosystem. arXiv [Preprint]. 2024. arXiv:2407.18322. doi:10.48550/arXiv.2407.18322.

10. 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]. Silver Spring (MD): FDA; 2025 Jan.

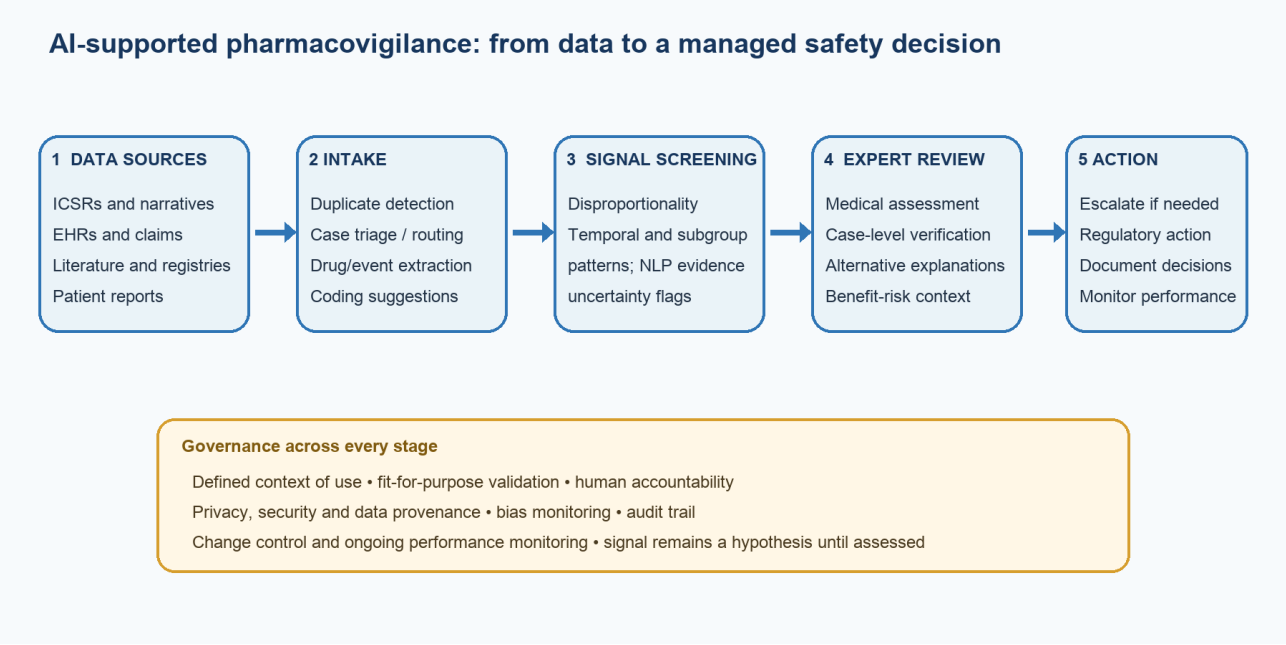
11. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). E2D(R1) guideline on post-approval safety data: definitions and standards for management and reporting of individual case safety reports. Geneva: ICH; 2025. Effective 2026 Mar 18.

12. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91(6):1010-1021. doi:10.1038/clpt.2012.50.

13. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform. 2015;54:202-212. doi:10.1016/j.jbi.2015.02.004.

14. Figures

Figure 1. Governed AI-supported workflow for adverse drug reaction detection and signal management. The illustration is conceptual and represents recommended process boundaries; it contains no experimental data.



Abbreviations: AI, artificial intelligence; EHR, electronic health record; ICSR, individual case safety report; NLP, natural language processing; PV, pharmacovigilance.