INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

Review Article

Integrating Artificial Intelligence and Natural Language Processing into Automated Adverse Event Signal Detection: A Comprehensive Review of Frameworks, Real-World Data, and Regulatory Challenges

Sakthikumar P*

Pachamuthu College of Pharmacy, Dharmapuri, Affiliated to The Tamil Nadu Dr. M.G.R. Medical University, Chennai

ARTICLE INFO

Published: 07 Aug 2026
Keywords:
pharmacovigilance; artificial intelligence; natural language processing; adverse event signal detection; disproportionality analysis; real-world data; regulatory science; machine learning.
DOI:
10.5281/zenodo.21843487

ABSTRACT

Pharmacovigilance, the science of detecting, assessing, and preventing adverse effects of medicines, has traditionally relied on spontaneous reporting systems and manually applied disproportionality statistics to generate safety signals. The explosive growth in the volume and diversity of drug-safety-relevant data -- spanning structured spontaneous report databases, electronic health records, clinical narratives, social media, and other real-world data sources -- has outpaced the practical capacity of manual and purely statistical signal-detection workflows. This review examines how artificial intelligence (AI) and natural language processing (NLP) are being integrated into automated adverse event signal detection, synthesising the current landscape of computational frameworks, real-world data sources, and the regulatory challenges that accompany their adoption. We describe the evolution from classical disproportionality methods -- the Proportional Reporting Ratio, Reporting Odds Ratio, Bayesian Confidence Propagation Neural Network, and Multi-item Gamma Poisson Shrinker -- toward machine-learning and deep-learning architectures capable of learning complex, non-linear associations across heterogeneous data. Particular attention is given to the role of transformer-based language models, including BioBERT, ClinicalBERT, PubMedBERT, and SpanBERT, in extracting adverse drug event mentions from unstructured clinical notes, biomedical literature, and social media text, and to hybrid frameworks that combine structured disproportionality analysis with unstructured-text mining and graph-based relational modelling. The review further surveys the principal real-world data infrastructures underpinning modern pharmacovigilance -- the FDA Adverse Event Reporting System, the WHO global individual case safety report database VigiBase, EudraVigilance, the FDA Sentinel Initiative, and the EMA's DARWIN EU platform -- and considers how AI-based tools are being layered onto these systems for federated querying, duplicate

***Corresponding Author:** Sakthikumar P
Address: *Pachamuthu College of Pharmacy, Dharmapuri Affiliated to The Tamil Nadu Dr. M.G.R. Medical University*
Email : sakthipsk22@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

detection, and case-processing automation. Finally, the review addresses the regulatory and governance challenges that currently constrain routine adoption of AI-augmented signal detection, including model transparency and explainability, validation and performance monitoring, algorithmic bias, data privacy, and the evolving guidance issued by the FDA, EMA, ICH, and CIOMS, concluding with a forward-looking assessment of the conditions under which AI/NLP-based signal detection is likely to become an accepted, auditable component of routine drug safety surveillance.

INTRODUCTION

Pharmacovigilance is formally defined as the science and set of activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other medicine-related problem, and it constitutes one of the central pillars of post-marketing drug safety oversight [1]. For most of its modern history, the discipline has depended on spontaneous reporting systems -- databases populated by voluntary reports from healthcare professionals, patients, and manufacturers -- combined with statistical disproportionality methods to flag drug-event combinations that are reported more frequently than would be expected by chance. This established workflow has been effective at identifying many important safety signals, but it was designed for an era in which the volume of safety-relevant data was orders of magnitude smaller than it is today.

The past decade has seen an unprecedented expansion in both the volume and heterogeneity of data that is potentially relevant to adverse event detection. Individual case safety report (ICSR) databases such as the United States Food and Drug Administration Adverse Event Reporting System (FAERS) and the World Health Organization's global database, VigiBase, now contain many millions of reports, while electronic health records (EHRs), clinical trial narratives, medical literature, and social media platforms represent vast, largely unstructured reservoirs of additional safety-relevant information [3]. Manually reviewing this volume of heterogeneous data is no longer practicable, and industry estimates suggest that traditional, largely manual case processing can consume up to two-thirds of a pharmacovigilance department's operating budget, creating strong pressure toward automation [7].

Artificial intelligence (AI) -- and, in particular, machine learning (ML), deep learning (DL), and natural language processing (NLP) -- has consequently emerged as a central focus of pharmacovigilance modernisation efforts. These technologies promise not merely to accelerate existing workflows but to enable qualitatively new capabilities: extraction of adverse event mentions directly from unstructured clinical notes and patient-generated text, integration of structured and unstructured data within unified signal-detection frameworks, and proactive, continuous surveillance across federated real-world data sources rather than periodic, batch-oriented review of spontaneous reports alone [1,2].

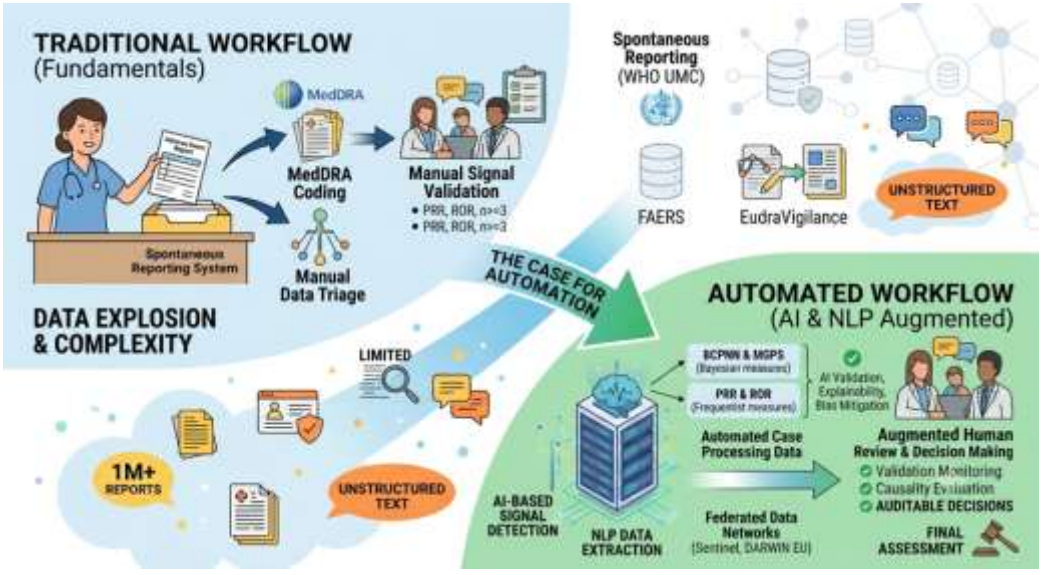
This review has three principal aims. First, it traces the evolution of adverse event signal detection from classical, purely statistical disproportionality methods toward AI- and NLP-augmented frameworks, describing the underlying computational approaches at a level of technical detail appropriate for pharmacy and pharmaceutical-sciences readers. Second, it surveys the principal real-world data infrastructures -- spontaneous reporting databases, EHR-linked surveillance networks, and social media -- that are now being mined using these methods, together with representative published applications. Third, it examines the regulatory, governance, and validation challenges that currently determine whether, and how, AI-augmented signal detection can be adopted within

regulated pharmacovigilance practice, drawing on the rapidly evolving guidance issued by the FDA, EMA, ICH, and CIOMS through 2025 and into 2026.

The review is organised as follows. Section 2 revisits the fundamentals of pharmacovigilance signal detection and the case for automation. Section 3 describes classical disproportionality analysis methods. Section 4 discusses the expansion of the signal-detection toolkit through machine learning. Section 5 introduces NLP foundations relevant to adverse event extraction. Section 6 examines transformer-based language models in this domain. Section 7 discusses hybrid and graph-based frameworks. Section 8 surveys real-world data infrastructures. Section 9 presents representative case-study applications. Section 10 addresses model validation and benchmarking. Section 11 discusses explainability and transparency. Section 12 reviews the regulatory landscape. Section 13 discusses data privacy, governance, and algorithmic bias. Section 14 addresses workflow integration and human-in-the-loop design. Section 15 offers future perspectives. Section 16 discusses the limitations of this review, and Section 17 concludes.

2. Pharmacovigilance Fundamentals and the Case for Automation

Signal management, as formally described in international guidance, comprises four sequential activities: signal detection, in which a potential drug-event association is first identified; signal validation, in which the evidence is assessed for plausibility and confirmed as warranting further evaluation; signal evaluation, in which the association is characterised in depth using available epidemiological and mechanistic evidence; and final assessment, in which a determination is made as to whether the signal constitutes a genuine, causally related adverse drug reaction [1]. Traditionally, signal detection has relied predominantly on disproportionality analysis of spontaneous report databases, while validation and evaluation have relied heavily on manual literature review, clinical judgement, and supplementary epidemiological studies.



The case for AI-based automation rests on several converging pressures. First, the sheer volume of ICSRs has grown dramatically; some large marketing authorisation holders now process well over one million individual case safety reports annually, a volume that strains manual triage and

coding capacity [7]. Second, an increasing proportion of potentially safety-relevant information exists in unstructured form -- free-text clinical narratives, physician notes, discharge summaries, call-centre transcripts, and social media posts -- that classical disproportionality methods, which operate on structured, coded data, cannot directly exploit [8]. Third, real-world data sources such as EHRs and administrative claims are increasingly viewed as complementary to spontaneous reports, offering denominator information and longitudinal patient histories that spontaneous reports typically lack, but their scale similarly precludes exhaustive manual review [3].

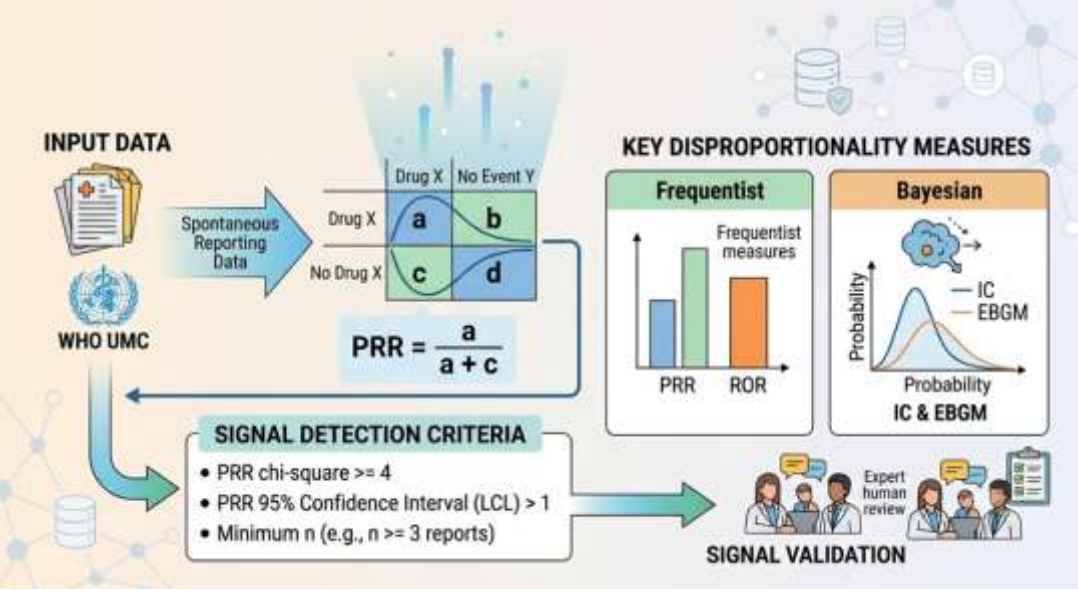
AI and NLP address these pressures in complementary ways. Machine-learning classifiers can learn complex, non-linear patterns across large structured datasets that may not be captured by fixed disproportionality formulae, while NLP techniques can extract structured adverse-event information from unstructured text, effectively converting previously inaccessible data into a form that downstream signal-detection algorithms can use [2,9]. Recent narrative and systematic reviews consistently report that AI-based tools are being applied across the full signal-management pipeline, though with substantially more maturity in automated case processing and signal detection than in signal validation, evaluation, and causality assessment, where methodological transparency remains mixed and gold-standard evaluation datasets are used inconsistently [1,4].

It is worth emphasising that automation in this context is not intended, nor is it generally proposed in the literature, as a wholesale replacement for the expert clinical and epidemiological judgement that underpins signal validation and causality assessment. Rather, the consistent framing across recent reviews is one of augmentation: AI and NLP tools are positioned to absorb the increasing volume and complexity of data at the detection and triage stages of the pipeline, thereby allowing qualified pharmacovigilance professionals to concentrate their expert judgement on the validation, evaluation, and causality-assessment stages where nuanced clinical and epidemiological reasoning remains, at present, difficult to fully automate [1].

3. Classical Disproportionality Analysis: The Statistical Foundation

Before considering how AI extends signal detection, it is necessary to understand the statistical methods it builds upon. Disproportionality analysis identifies drug-event combinations that are reported more frequently in a spontaneous report database than would be expected if the drug and the event were statistically independent, using the structure of a two-by-two contingency table constructed from report counts [11,19]. Four methods dominate current practice: the Proportional Reporting Ratio (PRR), the Reporting Odds Ratio (ROR), the Bayesian Confidence Propagation Neural Network (BCPNN), and the Multi-item Gamma Poisson Shrinker (MGPS) [11,18,20].

The PRR and ROR are frequentist measures, calculated directly from observed and expected report counts, and are valued for their computational simplicity and high sensitivity, though this comes at the cost of a higher false-positive rate, particularly for rare events with small report counts [11,19]. A drug-event combination is typically flagged as a signal under the PRR method when the chi-square statistic exceeds a defined threshold (commonly four) and the lower bound of the ninety-five percent confidence interval exceeds one, often combined with a minimum report count (commonly three) [14].



BCPNN and MGPS, by contrast, are Bayesian shrinkage methods, originally developed respectively at the WHO Uppsala Monitoring Centre and by DuMouchel for the FDA spontaneous reporting system [18,20]. These methods apply a Bayesian prior that shrinks disproportionality estimates toward the null for drug-event pairs with sparse data, improving specificity and stability for rare combinations at some cost in sensitivity relative to the frequentist measures [11,21]. In BCPNN, the key output statistic is the Information Component (IC), with a positive lower bound of its ninety-five percent credibility interval (commonly denoted IC025 greater than zero) taken as evidence of disproportional reporting; in MGPS, the analogous statistic is the Empirical Bayes Geometric Mean (EBGM), with its lower confidence bound (EBGM05) compared against a threshold, commonly two [14,16].

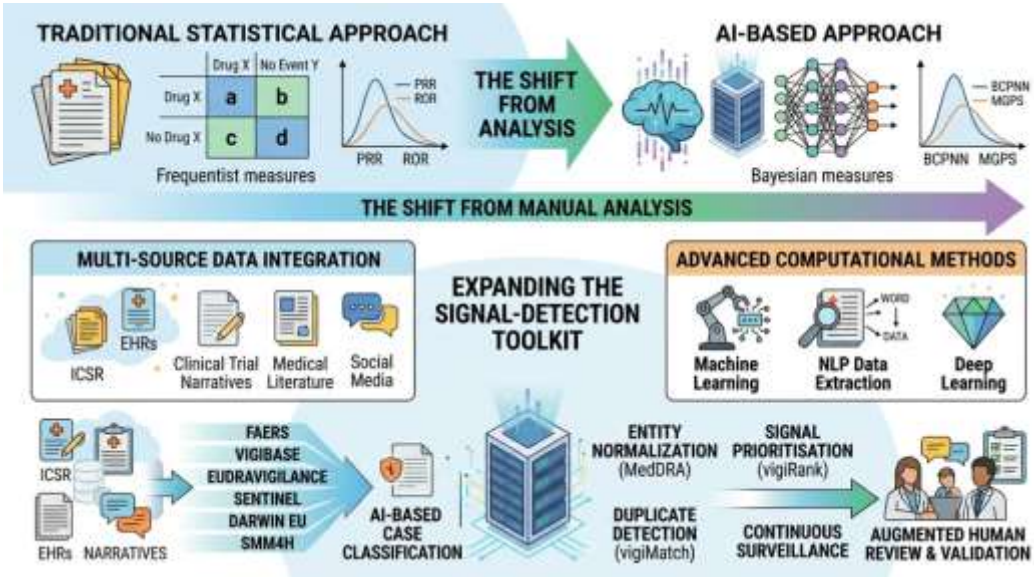
Because no single disproportionality method constitutes a universally accepted gold standard, contemporary pharmacovigilance studies increasingly report multiple measures in parallel and require concordance across methods -- for example, requiring that all four of PRR, ROR, BCPNN, and MGPS simultaneously indicate a signal -- before considering a drug-event association a validated candidate for further evaluation [11,16]. This multi-method convergence strategy, while conservative, has become a de facto standard in recently published disproportionality studies using FAERS and VigiBase data and represents an important methodological baseline against which AI-augmented signal-detection approaches are typically benchmarked [13,17].

Despite their continued centrality, classical disproportionality methods share several well-recognised limitations that motivate the move toward AI-based approaches: they operate exclusively on structured, MedDRA-coded fields and cannot directly exploit free-text narrative content; they are vulnerable to reporting biases, including stimulated reporting following media attention and systematic under-reporting of certain event types; and they generate purely statistical associations that require substantial downstream clinical judgement to distinguish genuine safety signals from confounding, indication bias, or coincidental reporting patterns [4,21].

4. From Statistics to Machine Learning: Expanding the Signal-Detection Toolkit

Machine-learning approaches to signal detection extend the classical disproportionality framework in several directions. Supervised classification models -- including logistic regression, random forests, and gradient boosting machines -- have been applied to predict the likelihood that a given case report, or a given drug-event pair, represents a genuine adverse drug reaction, using report-level features (patient demographics, concomitant medications, reporter type, narrative length) in addition to, or instead of, simple disproportionality statistics [1]. A recent critical evaluation of AI applications in signal management found that innovation was being driven particularly by high-performing ensemble methods such as random forest and gradient boosting, especially in the signal-detection stage of the pipeline, though methodological transparency across published studies remained inconsistent [1].

Beyond report-level classification, machine-learning models have been developed to predict adverse drug reactions directly from structured clinical data in hospitalised patients, leveraging electronic health record fields such as laboratory results, vital signs, and medication administration records to flag patients at elevated risk of an emerging adverse event before it is formally reported [2]. Such predictive approaches represent a shift from purely retrospective signal detection in aggregated spontaneous report data toward prospective, patient-level risk stratification, a distinction that has significant implications for how these tools are validated and regulated, as discussed further in Sections 10 and 12.



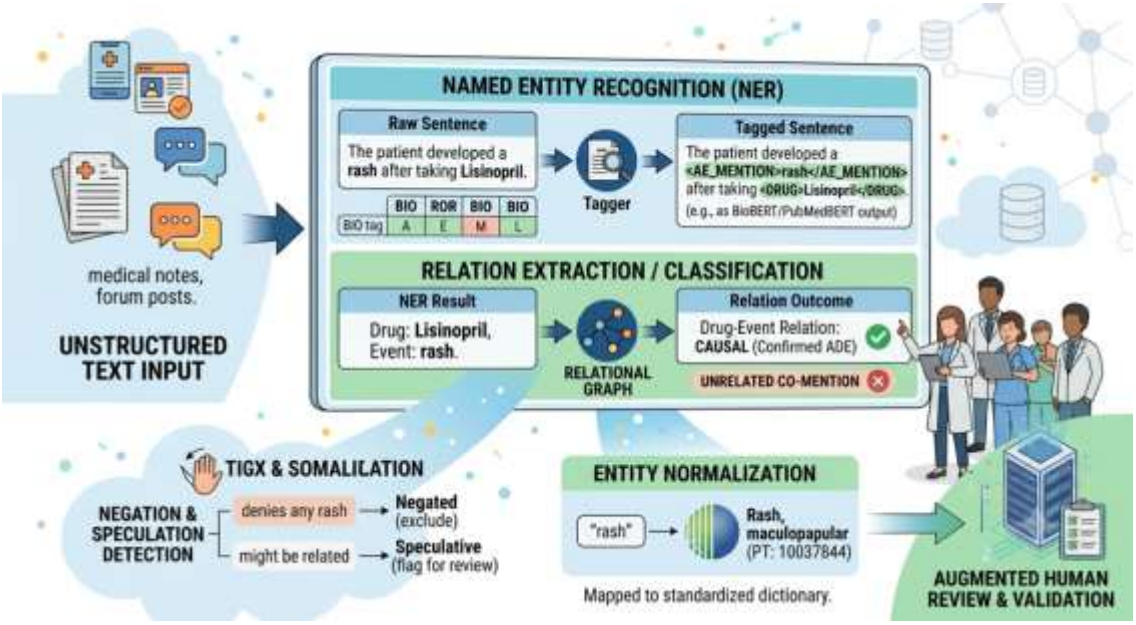
Graph-based and network approaches represent a further extension of the machine-learning toolkit. Because pharmacovigilance data can naturally be represented as a heterogeneous graph -- with nodes corresponding to patients, drugs, adverse events, and biological targets, and edges representing prescribing, co-occurrence, or interaction relationships -- graph neural networks (GNNs) have been proposed as a means of capturing complex, multi-way interactions, such as polypharmacy-driven adverse events, that are difficult to represent using conventional tabular machine-learning features [6]. Hybrid frameworks that combine structured demographic and laboratory data with unstructured clinical-note features, processed through deep-learning and

NLP pipelines, have been reported to outperform traditional disproportionality-only signal-detection methods while additionally providing more clinically interpretable, case-level predictions [5,6].

A recurring theme across this literature is that machine-learning models are rarely deployed as complete replacements for classical disproportionality analysis; rather, they are most often positioned as complementary tools -- for triage, prioritisation, or the incorporation of additional structured and unstructured data -- layered onto or run alongside the established PRR/ROR/BCPNN/MGPS framework described in Section 3 [1,4].

5. Natural Language Processing Foundations for Adverse Event Extraction

Natural language processing supplies the essential capability that distinguishes modern AI-augmented pharmacovigilance from classical disproportionality analysis: the ability to convert unstructured clinical or patient-generated text into structured, codeable adverse-event information. A comprehensive scoping review of supervised NLP methods for adverse drug event (ADE) detection in hospitalised-patient clinical narratives identified named entity recognition (NER) and relation extraction/classification as the two most frequent underlying tasks, accounting for the majority of the twenty-nine studies meeting inclusion criteria in that review [8].



Named entity recognition, in this context, involves identifying and labelling spans of text that correspond to specific entity types relevant to drug safety -- most commonly drug names, dosages, adverse event or symptom mentions, and temporal expressions -- typically framed computationally as a sequence-labelling problem in which each token in a sentence is classified according to a standard annotation scheme, such as the widely used Begin-Inside-Outside (BIO) tagging format [31,34]. Relation extraction builds on NER output to determine whether, and how, an identified drug mention and an identified adverse-event mention are causally or temporally related within the same clinical narrative, a step that is essential for distinguishing genuine adverse drug events from incidental co-mentions of a drug and a symptom that are unrelated to each other [15].

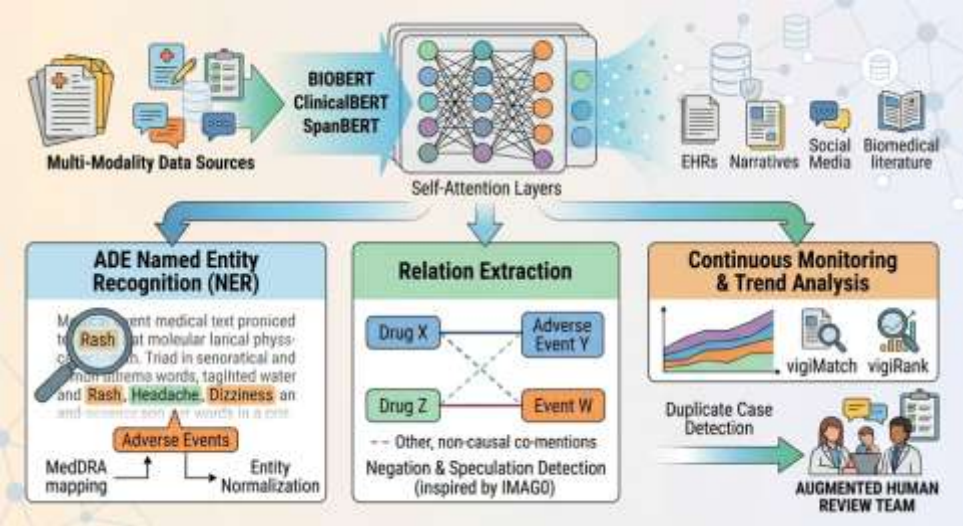
A substantial complicating factor in NLP-based ADE extraction, particularly for clinical narratives and social media text, is the frequent use of negation and speculative or hedged language -- for example, a clinical note stating that a patient "denies any rash" or a social media post noting that a symptom "might be" related to a medication. Naive entity-recognition systems that do not explicitly model negation and speculation risk substantially over-counting adverse event mentions, and dedicated negation-detection modules, ranging from rule-based algorithms such as NegEx to fine-tuned transformer-based classifiers, have been developed specifically to mitigate this source of error [17,35].

Early NLP systems for ADE extraction relied predominantly on rule-based pattern matching and classical statistical models such as conditional random fields, and portable, multi-corpus-trained text classifiers were an important step toward generalisable ADE detection across differing report styles and institutions [7]. The subsequent, and now dominant, generation of approaches is built on transformer-based contextual language models, which are discussed in detail in Section 6.

6. Transformer-Based Language Models in Pharmacovigilance

The introduction of transformer-based contextual language models has substantially advanced the state of the art in ADE extraction from both clinical and social media text. Domain-adapted models -- including BioBERT, pre-trained jointly on general-domain and biomedical-literature corpora [11], ClinicalBERT, pre-trained on clinical notes [10], and PubMedBERT, pre-trained from scratch on biomedical abstracts and full texts -- consistently outperform general-domain transformer models such as the original BERT architecture on biomedical and clinical NER benchmarks [29].

Comparative evaluations of transformer architectures for ADE extraction illustrate both the strength and the data-dependence of these models. In one systematic comparison of eleven transformer-based models evaluated on two standard ADE-extraction benchmarks -- the CADEC dataset, derived from patient forum posts, and the SMM4H dataset, derived from social media -- BioBERT achieved the strongest performance on the more clinically structured CADEC corpus, with an F1 score of approximately 86 percent, while SpanBERT, a model specifically designed for predicting text spans, achieved the best performance on the more linguistically diverse and informal SMM4H social media corpus, with an F1 score of approximately 84 percent [28,32]. This pattern -- in which no single model dominates across all text types -- underscores that model selection in pharmacovigilance NLP should be informed by the specific characteristics of the target text corpus, including its formality, domain specificity, and syntactic complexity [28].



The Social Media Mining for Health (SMM4H) shared task series, run annually since 2016 in conjunction with major computational linguistics venues, has played a particularly important role in benchmarking and advancing ADE-extraction NLP by providing standardised, annotated social media datasets and a competitive evaluation framework in which research groups worldwide submit and compare extraction systems [8,9]. Leading SMM4H submissions in recent years have consistently relied on fine-tuned transformer architectures, including BERT, SpanBERT, PubMedBERT, and task-specific variants such as EnDR-BERT, often combined with additional embedding or classification layers tailored to the noisy, informal characteristics of social media text [31,33].

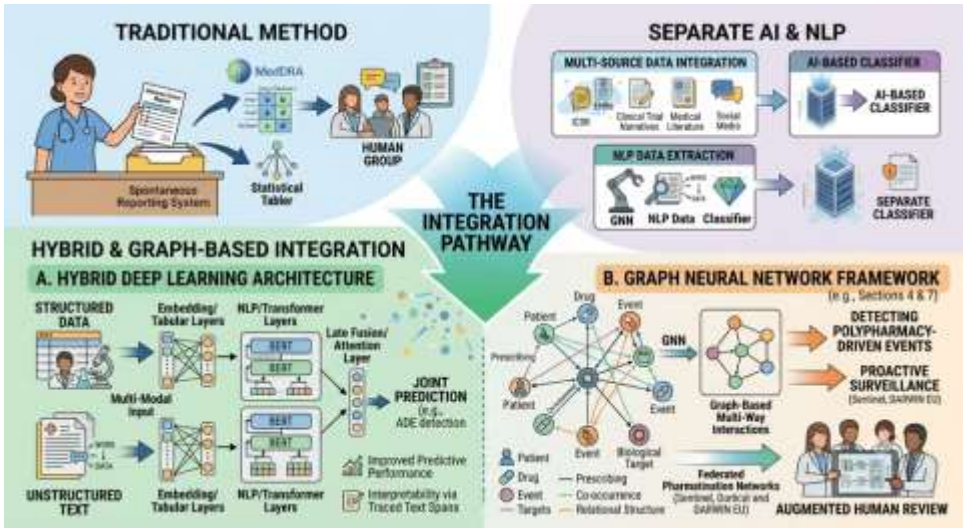
Beyond entity recognition, transformer models have been extended to the downstream task of entity normalisation -- mapping a free-text ADE mention (for example, a colloquial patient-reported symptom description) to a standardised MedDRA Preferred Term, a step that is essential for integrating NLP-extracted adverse events with existing coded pharmacovigilance databases and disproportionality pipelines. Multi-stage pipelines that first apply a fine-tuned transformer for entity recognition and subsequently apply a second transformer-based or zero-shot normalisation model to map recognised mentions to MedDRA terminology have achieved leading performance in recent SMM4H normalisation shared tasks, illustrating the increasing sophistication of end-to-end NLP pipelines for pharmacovigilance applications [30].

A further important methodological advance concerns robustness to negation and speculation, discussed conceptually in Section 5. Benchmark evaluations specifically designed to test ADE-extraction models under negated and speculative contexts have shown that even leading transformer-based extraction systems experience measurable performance degradation when evaluated on negation-heavy text, motivating the development of combined extraction-plus-negation-detection pipelines and dedicated robustness benchmarks such as the NADE benchmark [17,35].

7. Hybrid and Graph-Based Frameworks Integrating Structured and Unstructured Data

A significant methodological trend in recent pharmacovigilance research is the development of hybrid frameworks that integrate structured data (patient demographics, laboratory results,

concomitant medications) with unstructured data (clinical notes, narrative text) within a single predictive architecture, rather than treating disproportionality analysis and NLP-based text mining as separate, disconnected workflows [5,6]. Such hybrid approaches typically employ a deep-learning architecture in which structured features are processed through conventional tabular machine-learning or embedding layers, unstructured text is processed through an NLP or transformer-based encoder, and the resulting representations are fused -- through concatenation, attention mechanisms, or graph-based integration -- prior to a final prediction layer [5].



Graph neural networks warrant particular attention as an integrative framework. Because pharmacovigilance-relevant entities -- patients, drugs, adverse events, and, in some architectures, biological targets or pathways -- are naturally interconnected, representing this data as a heterogeneous graph and applying GNN architectures allows a model to learn from relational structure directly, for example capturing that two structurally unrelated drugs prescribed concomitantly are jointly associated with an adverse event via a shared node in the graph, a pattern that would be difficult to encode using standard tabular or sequence-based features alone [6]. Proposed GNN-based pharmacovigilance frameworks construct patient-drug-event graphs in which edges represent prescribing relationships or reported associations, and demonstrate improved detection of interactions arising from polypharmacy relative to methods that treat each drug-event pair independently [6].

Reported evaluations of these hybrid, multi-modal frameworks generally claim improved predictive performance relative to disproportionality-only or single-modality machine-learning baselines, alongside claims of improved interpretability arising from the ability to trace a prediction back to specific contributing structured features or text spans [5,6]. It should be noted, however, that such comparative claims are drawn predominantly from single-study evaluations using study-specific datasets and evaluation protocols, and broader, independent replication and benchmarking -- a challenge discussed further in Section 10 -- remains comparatively limited across this class of hybrid architecture.

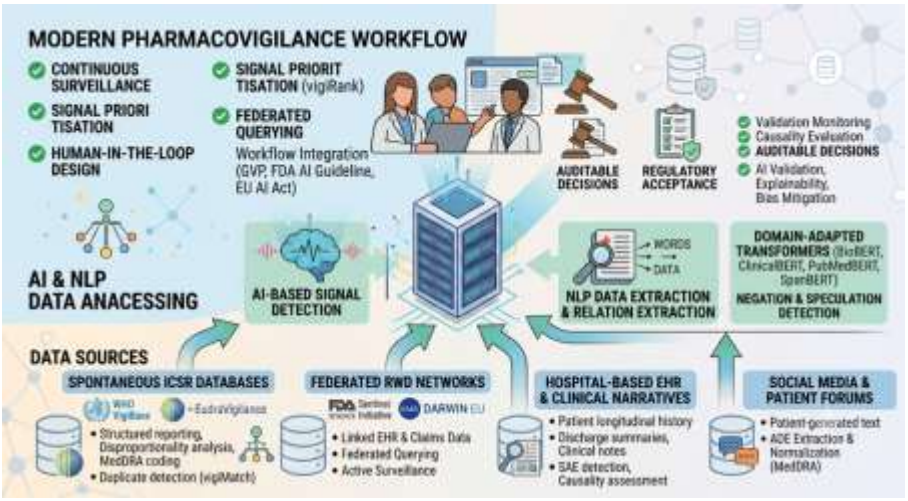
A related architectural consideration concerns how structured and unstructured representations are fused within a hybrid model. Early fusion approaches, in which raw or lightly processed structured and text-derived features are

concatenated prior to a shared downstream classifier, are architecturally simple but may struggle to capture cross-modal interactions -- for example, a specific combination of a laboratory abnormality and a particular phrase in a clinical note -- as effectively as late fusion or attention-based fusion approaches, in which each modality is first encoded separately and a learned attention mechanism determines how strongly each modality's representation should influence the final prediction for a given case. While the pharmacovigilance-specific literature on this architectural choice remains comparatively limited relative to the broader multi-modal machine-learning literature, the general finding from adjacent clinical-informatics domains -- that attention-based late fusion tends to outperform naive early fusion when the two modalities carry complementary rather than redundant information -- is likely to apply similarly to hybrid pharmacovigilance frameworks, and represents a promising area for further, pharmacovigilance-specific methodological investigation.

8. Real-World Data Infrastructures Underpinning AI-Augmented Pharmacovigilance

8.1 Spontaneous reporting databases. The FDA Adverse Event Reporting System (FAERS) and the WHO's global individual case safety report database, VigiBase, maintained by the Uppsala Monitoring Centre, remain the primary structured data sources for disproportionality-based signal detection, together containing tens of millions of case reports spanning decades of post-marketing experience [12,16]. FAERS in particular has become a favoured resource for AI- and machine-learning-augmented pharmacovigilance research owing to its public accessibility and standardised MedDRA and ATC coding, with recent published studies analysing FAERS datasets spanning the full period from the early 2000s through 2024 and beyond [13,14,16].

8.2 Regional and regulatory-network systems. The European Union's EudraVigilance system serves an analogous function within the European regulatory network, receiving individual case safety reports from marketing authorisation holders under strict timetables specified in EU pharmacovigilance legislation and supporting GVP Module-based signal-detection activity by the EMA and national competent authorities [23]. Notably, the Uppsala Monitoring Centre has already deployed machine-learning-based tools -- including vigiMatch, for duplicate-report detection, and vigiRank, for signal prioritisation -- directly within VigiBase's operational workflow, illustrating that AI adoption within core spontaneous-reporting infrastructure is not merely prospective but is already an operational reality at some pharmacovigilance centres [22].



8.3 Federated real-world data networks. Beyond spontaneous reporting, real-world data networks built on electronic health records and administrative claims have become an increasingly important complementary data source. The FDA's Sentinel Initiative aggregates real-world healthcare data across multiple partner organisations to support active, ongoing safety surveillance and has reportedly been used for several hundred discrete safety analyses since its inception [20]. The EMA's analogous initiative, DARWIN EU (Data Analysis and Real World Interrogation Network), similarly provides federated access to real-world healthcare data across the European Union to support regulatory decision-making, and recent industry analyses describe both Sentinel and DARWIN EU as increasingly "AI-augmented at the routine-monitoring layer," combining federated querying across harmonised, OMOP-common-data-model-shaped cohorts with NLP-based extraction from linked clinical notes [20].

8.4 Electronic health records and clinical narratives. Within individual healthcare systems, electronic health records represent a rich but heterogeneous data source for AI-based adverse event detection, combining structured fields (diagnoses, laboratory results, medication administration records) with extensive unstructured clinical narrative content. Systematic reviews of AI-based models in pharmacoepidemiology note that the large majority of published studies to date have relied primarily on structured EHR or administrative-claims fields, with only a minority incorporating NLP components to exploit narrative text, and none, in one recent systematic review's sample, combining EHR-based modelling with spontaneous-report data in a single, unified framework -- an observation that highlights a significant, currently unrealised opportunity for more fully integrated, multi-source AI pharmacovigilance frameworks [3].

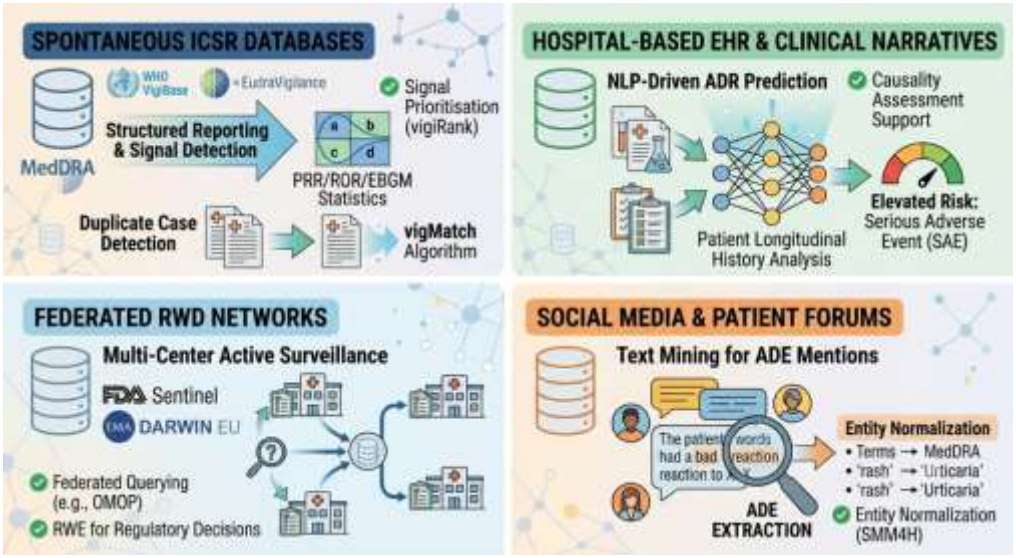
8.5 Social media and patient-generated content. Social media platforms, patient forums, and other patient-generated text sources have emerged as a distinctive real-world data modality for pharmacovigilance, offering the potential to capture patient-reported adverse events -- including those a patient might not think significant enough to report to a healthcare provider or formal reporting system -- at a scale and immediacy not achievable through traditional channels [28]. The SMM4H shared task series, discussed in Section 6, has been instrumental in developing and benchmarking the NLP methods needed to extract reliable, normalised adverse-event information from this inherently noisy,

informal, and linguistically diverse data source [8,9].

9. Representative Applications Across Data Modalities

9.1 Hospital-based adverse drug reaction prediction. Machine-learning models trained on discharge-summary and clinical-narrative data have been used to predict adverse drug reactions in hospitalised patients, with one representative study reporting a model capable of predicting several hundred distinct adverse drug events from discharge summaries, a substantial proportion of which were subsequently clinically validated, illustrating the potential of such models to surface adverse events that would otherwise remain undetected under spontaneous-reporting-only surveillance [2].

9.2 Large-scale FAERS disproportionality studies. A large and rapidly growing body of published pharmacovigilance research applies the classical disproportionality methods described in Section 3 -- typically two to four of PRR, ROR, BCPNN, and MGPS/EBGM in combination -- to FAERS and VigiBase data to characterise the adverse-event profile of specific drugs or drug classes, ranging from neurotoxicity signals associated with carbapenem antibiotics to hepatobiliary toxicity associated with immune checkpoint inhibitors and ocular adverse events associated with targeted anti-cancer therapies [12,13,18]. While these studies do not always incorporate AI or NLP components directly, they constitute the empirical and methodological substrate against which AI-augmented signal-detection tools are most frequently benchmarked, and an increasing number now supplement disproportionality analysis with machine-learning-based case classification or NLP-based narrative review [16].



9.3 Social-media-derived pharmacovigilance signals. Transformer-based NLP pipelines applied to social media text, benchmarked through the SMM4H shared task series, have demonstrated the feasibility of extracting clinically meaningful, MedDRA-normalised adverse-event mentions directly from informal, patient-generated posts, with leading systems achieving entity-recognition F1 scores in the low-to-mid eighty percent range on standard social media benchmarks [28,30]. Although social-media-derived signals require careful validation before being treated as equivalent in evidentiary weight to structured spontaneous reports, they offer a genuinely

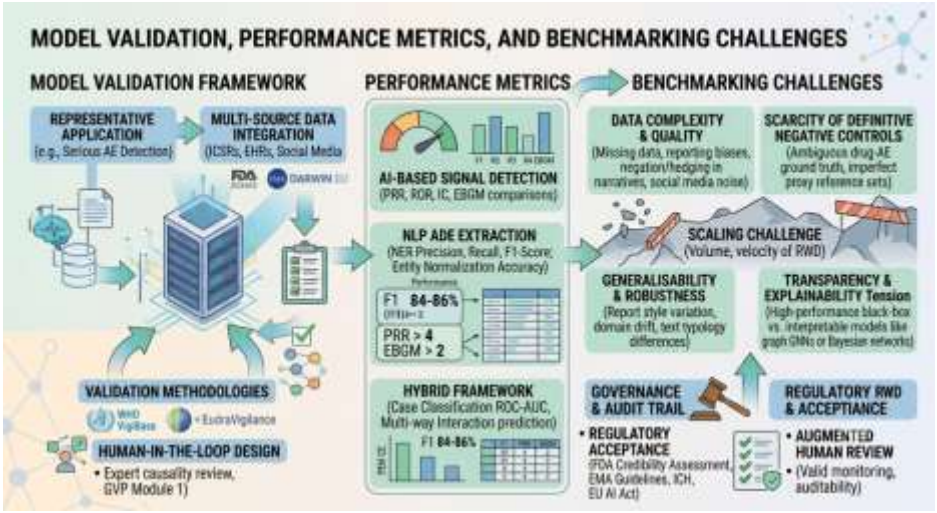
complementary early-warning capability, particularly for adverse events that patients experience as bothersome but do not consider severe enough to report through formal channels.

9.4 Serious adverse event detection in clinical trial data. Hybrid deep-learning and NLP frameworks combining structured clinical-trial data fields with unstructured clinical-note content have been proposed specifically for the detection of serious adverse events (SAEs) during clinical trials, an application domain with particularly high stakes given the direct relevance of SAE detection to trial-participant safety and to Good Clinical Practice compliance, with reported frameworks claiming improved detection accuracy and interpretability relative to conventional, single-modality signal-detection approaches [5,6].

9.5 Duplicate detection and case-processing automation. A distinct but complementary application of AI within pharmacovigilance concerns the automation of routine case-processing tasks that precede formal signal detection, including duplicate-report identification, case triage, and structured-field extraction from incoming case narratives. The Uppsala Monitoring Centre's operational deployment of the vigiMatch algorithm for probabilistic duplicate detection within VigiBase, and comparable proprietary tools reported by individual marketing authorisation holders for automated case intake, illustrate that AI adoption in this "upstream" portion of the pharmacovigilance workflow has, in several organisations, already progressed beyond the pilot stage into sustained operational use, in contrast to the more nascent state of AI adoption in signal validation and causality assessment discussed in Sections 10 and 11 [7,22].

10. Model Validation, Performance Metrics, and Benchmarking Challenges

Rigorous validation of AI-based signal-detection tools presents challenges that differ substantially from those associated with classical disproportionality methods. Because AI models are trained on finite datasets and may overfit to idiosyncrasies of their training corpus, their performance must be assessed on held-out or externally sourced test data, using standard classification metrics -- precision, recall, and the F1 score for entity-level extraction tasks, and measures such as area under the receiver operating characteristic curve for report- or patient-level risk-prediction tasks. A critical evaluation of published AI pharmacovigilance signal-management studies found that only a subset employed publicly available, "gold standard" positive and negative control datasets, and that methodological transparency regarding validation procedures was mixed across the literature, complicating cross-study comparison [1].



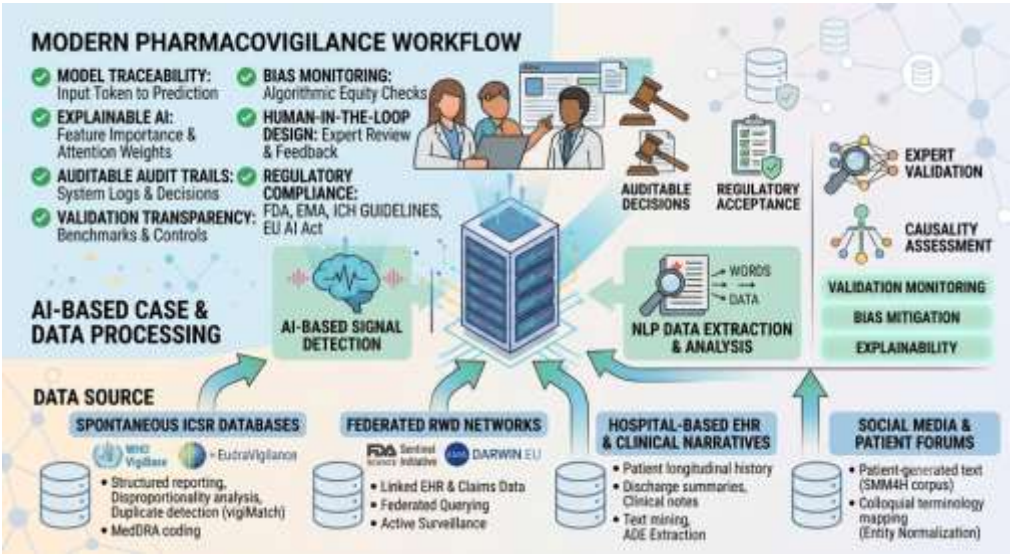
Benchmark datasets such as CADEC (derived from patient-authored forum posts describing adverse drug events) and the annually refreshed SMM4H shared-task corpora have played a valuable role in providing standardised evaluation conditions for NLP-based ADE-extraction models, allowing direct, controlled comparison of competing architectures under identical data and annotation conditions [28,32]. Nonetheless, performance on these benchmarks does not automatically generalise to real-world deployment conditions, since production data streams may differ from benchmark corpora in report style, therapeutic area coverage, temporal drift in language use, and the prevalence of negation or speculative language, all of which have been shown to degrade extraction performance relative to benchmark conditions [17,35].

A further validation challenge specific to pharmacovigilance concerns the scarcity of definitive negative controls: because the ground truth for whether a given drug genuinely causes a given adverse event is frequently uncertain even after extensive epidemiological investigation, evaluating a signal-detection model's true precision and recall against real-world outcomes is intrinsically more difficult than in many other applied machine-learning domains, and published evaluations often rely on established reference sets of previously confirmed drug-event associations as an imperfect proxy for ground truth [1,4]. This uncertainty is compounded when incomplete or systematically missing data -- for example, missing indication, dose, or concomitant medication fields in spontaneous reports -- biases both the training and evaluation of AI models in ways that are difficult to fully characterise or correct [4].

11. Explainability and Transparency in AI-Driven Signal Detection

The relatively opaque, "black-box" character of many high-performing machine-learning and deep-learning architectures -- particularly deep neural networks and large transformer-based language models -- poses a distinctive challenge for pharmacovigilance applications, where downstream decisions can affect regulatory action, prescribing practice, and patient safety, and where clinical and regulatory reviewers must be able to understand and justify the basis for a flagged signal [1]. Recent critical reviews of AI in pharmacovigilance signal management explicitly identify the level of methodological transparency as a key differentiator between published studies, and recommend that future work make

explainability an explicit design requirement rather than an afterthought [1].



Several categories of explainability technique have been applied or proposed in this domain. Post-hoc, model-agnostic explanation methods -- such as attention-weight visualisation for transformer-based NLP models, or feature-importance techniques for tabular machine-learning classifiers -- can indicate which input tokens, features, or clinical variables contributed most strongly to a given model output, providing a partial, though not fully causal, account of model behaviour. Hybrid frameworks that explicitly combine structured feature-based prediction with traceable NLP-extracted text spans have been reported to offer improved interpretability relative to end-to-end deep-learning models, precisely because a flagged signal can be traced back to a specific, human-readable passage of clinical text rather than an opaque internal representation [5,6].

Bayesian network approaches, in which causal or probabilistic relationships between variables are represented as an explicit, human-interpretable graphical structure, have also been proposed and piloted as an alternative to less transparent deep-learning architectures specifically because their structure can be directly reviewed, audited, and, where necessary, adjusted by domain experts, a property considered particularly valuable for causality-assessment tasks in operational pharmacovigilance centres. This preference for inherently interpretable model classes in high-stakes, regulator-facing applications reflects a broader and increasingly consequential tension in the field between the highest-performing AI architectures and the practical requirement for auditable, explainable decision support, a tension that features prominently in the regulatory guidance discussed in Section 12.

12. The Regulatory Landscape: FDA, EMA, ICH, and CIOMS Guidance

Regulatory engagement with AI in pharmacovigilance has accelerated markedly since 2024. In January 2025, the FDA released draft guidance titled "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products," which introduces a risk-based credibility assessment framework requiring sponsors to establish and justify the credibility of an AI model for its specific context of use, with the

degree of required evidence scaled to the model's influence on regulatory decisions and the risk associated with the decision it supports [21,24].

The EMA has pursued a parallel, though independently developed, regulatory track. In September 2023 the EMA published a draft reflection paper on the use of artificial intelligence across the medicinal product lifecycle, which was finalised in 2024; the finalised paper emphasises general principles -- data traceability, risk-based governance, and cross-stakeholder collaboration -- rather than prescriptive technical rules, and explicitly extends to AI applications within pharmacovigilance, alongside AI use in clinical development and manufacturing [23,24]. The EMA has concurrently pursued concrete internal initiatives, including a dedicated AI workplan and industry-platform working group tasked with establishing guiding principles for AI in pharmacovigilance specifically, organised around the themes of governance and accountability, risk-based approach, human oversight, validity and robustness, transparency and explainability, data privacy, and fairness and equity [25].

A particularly significant recent development is the joint publication, by the FDA and EMA together, of ten guiding principles for the use of AI across the medicines lifecycle, reported as published in January 2026, representing an unusual degree of direct transatlantic regulatory alignment on AI governance in the pharmaceutical sector [26,27]. This joint publication follows the EMA and FDA's earlier, separately developed guidance documents and signals an intent to harmonise expectations for evidence generation, monitoring, and risk-based oversight of AI tools across the product lifecycle, including pharmacovigilance applications, rather than allowing divergent regional standards to emerge [26].

Beyond the FDA and EMA, the International Council for Harmonisation (ICH) and the Council for International Organizations of Medical Sciences (CIOMS) have also engaged with AI governance in pharmacovigilance, with CIOMS Working Group XIV specifically addressing AI applications in pharmacovigilance signal management practice, reflecting a broader international convergence of interest, even as detailed, binding technical standards for AI validation in this domain remain, as of this writing, still under active development at most agencies [7].

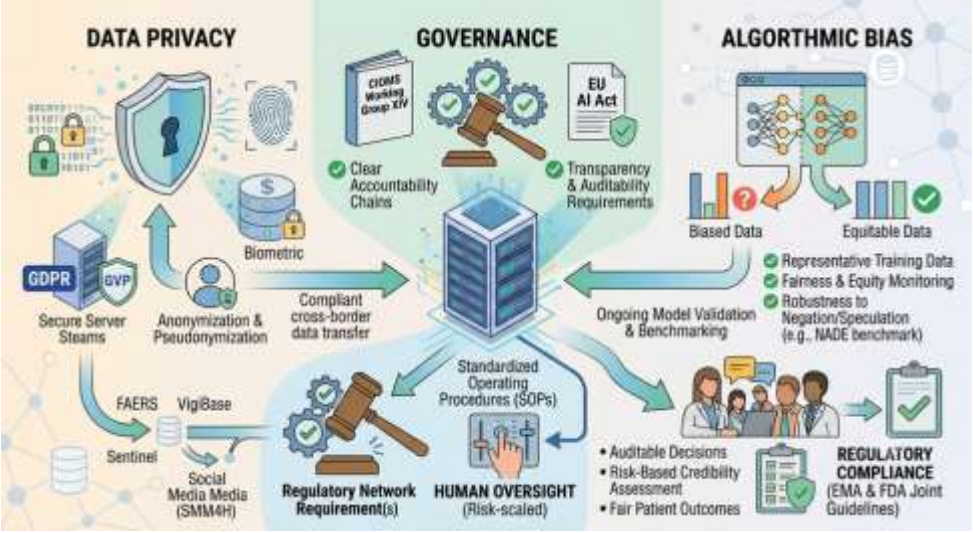
Within the European Union specifically, AI applications in pharmacovigilance are also subject to the broader EU Artificial Intelligence Act, whose obligations for general-purpose AI models began phasing in from August 2025, with more stringent obligations for AI systems classified as "high-risk" -- a category that plausibly captures at least some AI applications used in drug safety decision-making -- beginning to phase in from August 2026 onward, adding a further, cross-cutting layer of regulatory requirement beyond sector-specific pharmacovigilance guidance [26].

Taken together, these parallel and increasingly converging regulatory tracks -- FDA risk-based credibility assessment, the EMA's reflection paper and pharmacovigilance-specific guiding principles, joint FDA-EMA lifecycle principles, CIOMS Working Group XIV, and the EU AI Act -- indicate that the regulatory environment for AI in pharmacovigilance, while still under active development, is moving distinctly toward risk-proportionate, transparency-oriented governance rather than either blanket prohibition or unregulated adoption. For pharmaceutical companies, technology developers, and pharmacovigilance centres, this trajectory implies that early, proactive engagement with emerging guidance -- rather than a wait-and-see posture -- is likely to be the more strategically sound approach to AI adoption in this domain.

13. Data Privacy, Governance, and Algorithmic Bias

AI-augmented pharmacovigilance systems that draw on electronic health records, clinical narratives, or social media inevitably raise data-privacy considerations that extend beyond those associated with traditional, already-de-identified spontaneous report databases. Within the European Union, any AI tool that processes patient-level clinical data, including cross-border data used to train or validate multinational pharmacovigilance models, must comply with the General Data Protection Regulation (GDPR) and applicable national law, and must appropriately interface with existing regulatory data infrastructure such as EudraVigilance without compromising these obligations [22].

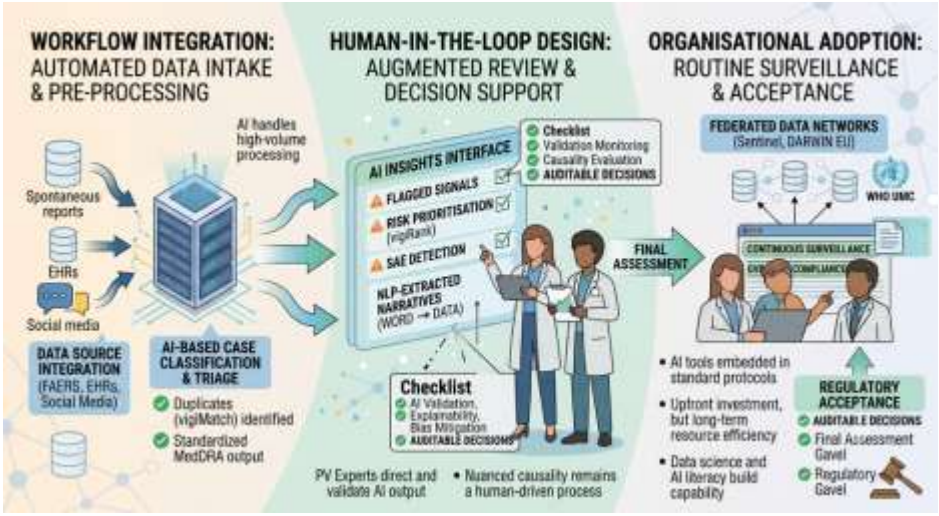
Algorithmic bias represents a further, closely related governance concern. Because AI models learn from historical data, they risk perpetuating or amplifying pre-existing biases present in that data -- for example, systematic under-representation of certain demographic groups in clinical trial and EHR datasets, or differential reporting behaviour by healthcare-system type or geographic region in spontaneous report databases -- with the result that a model's signal-detection sensitivity may vary, in ways that are not always transparent, across patient subgroups [1]. Recent critical commentary specifically highlights the risk that AI-based signal detection and causality assessment, if built on incomplete or systematically non-representative data, may generate biased insights that are mistaken for objective, data-driven conclusions precisely because of the apparent rigor conferred by computational methods .



Regulatory guiding principles emerging from the EMA's pharmacovigilance-specific AI workstream explicitly name data privacy and fairness/equity as core governance objectives, alongside governance and accountability, risk-based approach, human oversight, and validity and robustness, reflecting a broadly shared view among stakeholders -- spanning regulators, industry, and academic pharmacovigilance centres -- that technical performance metrics alone are insufficient grounds for approving an AI tool for use in regulated drug-safety decision-making [25]. Ensuring adequate human oversight in particular has emerged as a recurring theme: regulatory guidance across jurisdictions consistently frames AI as intended to augment, rather than replace, expert human judgement in signal validation, evaluation, and causality assessment, with the appropriate degree of human oversight scaled to the risk and consequence of the specific decision the AI tool supports [21,25].

14. Workflow Integration, Human-in-the-Loop Design, and Organisational Adoption

Even a technically validated, adequately transparent AI model delivers limited practical benefit unless it can be integrated into existing pharmacovigilance workflows in a manner that pharmacovigilance professionals can trust and productively use. Published accounts of practical AI implementation in operational pharmacovigilance centres emphasise that successful adoption typically follows a human-in-the-loop design pattern, in which AI or NLP tools are used to triage, prioritise, or pre-process large volumes of data -- for example, flagging candidate signals or extracting draft structured fields from narrative text -- while final validation, causality assessment, and regulatory reporting decisions remain under the explicit control of qualified pharmacovigilance personnel.



Organisational adoption barriers identified across the recent literature include the need for substantial upfront investment in computational infrastructure and data-integration capability; the requirement for ongoing model validation, monitoring, and retraining as underlying data distributions and clinical documentation practices evolve over time; and workforce considerations, including the need to build sufficient data-science and AI-literacy capability within pharmacovigilance teams that have traditionally been staffed predominantly by clinical and regulatory-science professionals [1]. A recent review focused specifically on the transition of AI in pharmacovigilance "from experimental applications... to being considered for routine use" identifies practical implementation challenges -- consistent and transparent AI performance over time, reduction of multiple sources of bias, and interpretability -- as the primary obstacles to this transition, rather than any remaining fundamental limitation in underlying AI or NLP technical capability.

Successful case studies of AI adoption in operational pharmacovigilance settings, such as the deployment of an expert-defined Bayesian network tool for causality assessment within a pharmacovigilance centre, illustrate a broader principle: that AI tools designed with explicit input from, and interpretability for, domain experts, and positioned to support rather than supplant expert judgement, are more readily integrated into regulated workflows than technically more powerful but less transparent alternatives. This principle is likely to remain central to organisational AI adoption strategies in pharmacovigilance for the foreseeable future, given the consistent emphasis, across the regulatory guidance reviewed in Section 12, on human oversight as a non-negotiable governance requirement.

15. Future Perspectives

Several trends are likely to shape the continued integration of AI and NLP into adverse event signal detection over the coming years. First, the trajectory of regulatory convergence exemplified by the joint FDA-EMA guiding principles suggests that pharmaceutical companies and pharmacovigilance technology developers may increasingly be able to design AI validation and governance processes against a single, internationally harmonised standard, rather than navigating substantially divergent regional requirements, which would meaningfully lower

the practical barrier to routine adoption of AI-augmented signal-detection tools [26,27].

Second, the current systematic-review evidence base indicates that the large majority of published AI pharmacoepidemiology studies to date rely on a single data modality -- most commonly structured EHR or claims data -- with comparatively few studies integrating spontaneous-report, EHR, and unstructured-text data within a single, unified analytic framework [3]. Closing this integration gap, through the kind of hybrid and graph-based architectures discussed in Section 7, represents one of the more promising and currently under-realised directions for methodological advancement, with the potential to meaningfully improve both sensitivity and specificity relative to single-modality approaches.

Third, continued refinement of explainable and inherently interpretable AI architectures -- including hybrid models that preserve traceability to specific structured features or text spans, and probabilistic graphical approaches such as Bayesian networks -- is likely to remain a priority, given the sustained regulatory and professional emphasis on transparency, auditability, and human oversight discussed in Sections 11 through 14. Fourth, as large language models continue to advance in general capability, their application to pharmacovigilance-specific tasks -- including automated literature surveillance, structured case-narrative summarisation, and natural-language query interfaces over federated real-world data networks -- is likely to expand substantially beyond the entity-recognition and classification tasks that have dominated the field to date, provided that appropriate validation, human oversight, and governance frameworks, consistent with emerging regulatory expectations, are established in parallel with this expanded functional scope [20,26].

Finally, addressing the data-quality, completeness, and representativeness challenges that currently limit both classical disproportionality analysis and AI-based approaches alike -- including systematically missing report fields, demographic under-representation, and reporting biases -- will likely require coordinated action extending beyond individual AI model development, encompassing improvements to spontaneous-reporting-system design, incentives for more complete case documentation, and continued expansion of federated real-world data infrastructure such as Sentinel and DARWIN EU [3].

16. Limitations of This Review

As a narrative review rather than a systematic review conducted under a pre-registered protocol, this work does not claim to comprehensively enumerate every published study of AI or NLP applications in adverse event signal detection, and the selection of illustrative studies and frameworks, while informed by recent systematic reviews and widely cited primary literature, inevitably reflects the authors' judgement as to which developments are most representative and pedagogically useful for a pharmaceutical-sciences readership.

The field itself is evolving rapidly, and several of the regulatory developments discussed in Section 12 -- including the joint FDA-EMA guiding principles and the phased implementation of EU AI Act obligations -- were, at the time of writing, either very recently finalised or still subject to further elaboration; readers should verify the current status of any specific regulatory requirement directly against the issuing authority's published guidance before relying on it for compliance purposes. Similarly, quantitative performance figures cited from individual studies (for example, specific F1 scores for named entity

recognition on particular benchmark datasets) reflect the reported results of the cited studies under their specific experimental conditions and should not be interpreted as universally generalisable performance guarantees for the models or architectures named.

Finally, because this review draws substantially on recently published systematic reviews and narrative reviews rather than exclusively on primary research articles, it necessarily inherits some of the scope and selection choices of those secondary sources; where possible, this has been made explicit in the text through direct attribution to the citing review rather than implying independent verification of every underlying primary finding.

CONCLUSION

Artificial intelligence and natural language processing are transforming adverse event signal detection from a largely manual, spontaneous-report-centred discipline into an increasingly automated, multi-source analytic capability. Classical disproportionality methods -- PRR, ROR, BCPNN, and MGPS -- remain the statistical backbone of routine signal detection and continue to serve as the benchmark against which newer approaches are evaluated, but they are being substantially extended by machine-learning classifiers, transformer-based NLP models capable of extracting adverse-event information directly from unstructured clinical and social media text, and hybrid, graph-based frameworks capable of integrating structured and unstructured data within a unified analytic architecture.

This transformation is occurring against a backdrop of real-world data infrastructure -- FAERS, VigiBase, EudraVigilance, the FDA Sentinel Initiative, and the EMA's DARWIN EU network -- that is itself expanding in scale and, in some cases, already incorporating AI-based tools directly into operational workflows. At the same time, the evidence reviewed in this paper indicates that the transition from experimental AI applications to fully routine, regulator-accepted use remains constrained less by fundamental technical limitations than by practical challenges of validation, explainability, bias mitigation, and governance -- challenges that regulatory bodies including the FDA, EMA, ICH, and CIOMS are now addressing with increasing urgency and, encouragingly, increasing international coordination.

For the field to fully realise the potential of AI- and NLP-augmented signal detection, continued attention will be needed on several fronts simultaneously: rigorous, transparent validation using appropriate benchmark and gold-standard datasets; explicit design for explainability and human oversight rather than treating these as post-hoc additions; proactive attention to algorithmic bias and data representativeness; and close, sustained engagement between AI developers, pharmacovigilance practitioners, and regulatory authorities as governance frameworks continue to mature. Laboratories, regulatory scientists, and pharmacovigilance professionals who engage with these technologies as complementary, auditable extensions of established pharmacovigilance science -- rather than as opaque replacements for expert judgement -- are best positioned to realise their benefits for patient safety while maintaining the rigor and accountability that drug safety surveillance demands.

Table 1. Classical disproportionality methods used in spontaneous-report signal detection

Method	Type	Key Statistic	Typical Signal Threshold
PRR	Frequentist	Proportional Reporting Ratio	PRR chi-square >= 4, lower 95% CI > 1, n >= 3 reports
ROR	Frequentist	Reporting Odds Ratio	Lower 95% CI of ROR > 1
BCPNN	Bayesian	Information Component (IC)	IC025 (lower 95% credibility bound) > 0
MGPS	Bayesian	Empirical Bayes Geometric Mean (EBGM)	EBGM05 (lower 95% CI) > 2

Table 2. Principal real-world data infrastructures relevant to AI/NLP-augmented pharmacovigilance

Data Source	Type	Primary Role in AI/NLP Signal Detection
FAERS	Spontaneous ICSR database (USA)	Structured disproportionality analysis; AI/ML case classification
VigiBase	Global spontaneous ICSR database (WHO-UMC)	Disproportionality analysis; ML-based duplicate detection (vigiMatch) and signal prioritisation (vigiRank)
EudraVigilance	EU regional ICSR database	GVP-aligned signal detection; emerging AI/NLP pilots
FDA Sentinel Initiative	Federated EHR/claims network (USA)	Active surveillance; federated querying; NLP extraction from linked notes
EMA DARWIN EU	Federated real-world data network (EU)	Regulatory-grade real-world evidence generation; AI-augmented cohort analysis
Social media / patient forums	Unstructured patient-generated text	NLP-based ADE extraction and normalisation (e.g., SMM4H benchmarks)

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
NLP	Natural Language Processing
ML / DL	Machine Learning / Deep Learning
ADE / ADR	Adverse Drug Event / Adverse Drug Reaction
ICSR	Individual Case Safety Report
FAERS	FDA Adverse Event Reporting System
VigiBase	WHO Global Individual Case Safety Report Database
EudraVigilance	EU Pharmacovigilance Database
DARWIN EU	Data Analysis and Real World Interrogation Network (EMA)
MedDRA	Medical Dictionary for Regulatory Activities
PRR / ROR	Proportional Reporting Ratio / Reporting Odds Ratio
BCPNN	Bayesian Confidence Propagation Neural Network
MGPS / EBGM	Multi-item Gamma Poisson Shrinker / Empirical Bayes Geometric Mean
NER	Named Entity Recognition
BERT	Bidirectional Encoder Representations from Transformers
GNN	Graph Neural Network
EHR	Electronic Health Record
SMM4H	Social Media Mining for Health (shared task series)
CIOMS	Council for International Organizations of Medical Sciences
ICH	International Council for Harmonisation
GVP	Good Pharmacovigilance Practices
EU AI Act	European Union Artificial Intelligence Act

Glossary of Key Terms

Term	Definition
Signal detection	The activity of identifying a potential, previously unrecognised or incompletely characterised drug-adverse event association from available data, warranting further evaluation.

Disproportionality analysis	A family of statistical methods that flag drug-event combinations reported more frequently than expected under an assumption of independence, using structured spontaneous report data.
Named entity recognition (NER)	An NLP task that identifies and labels spans of text corresponding to predefined categories, such as drug names or adverse event mentions.
Transformer model	A neural network architecture based on self-attention mechanisms, underlying contextual language models such as BERT and its biomedical/clinical variants.
Real-world data (RWD)	Data relating to patient health status or care delivery collected outside the context of randomised controlled trials, including EHRs, claims data, and registries.
Human-in-the-loop	A system design in which AI or automated outputs are reviewed, validated, or overridden by a qualified human expert before being acted upon.
Explainability	The degree to which the internal reasoning or decision basis of a model can be understood, audited, and communicated to a human reviewer.
Federated data network	An infrastructure that allows a query or analysis to be run across multiple, separately governed data sources without centralising the underlying patient-level data.

REFERENCES

1. Warner J, Prada Jardim A, Albera C, et al. Artificial Intelligence: Applications in Pharmacovigilance Signal Management. Drug Saf. 2025;39(3):183-198.
2. Artificial intelligence (AI) in pharmacovigilance: a systematic review on predicting adverse drug reactions (ADR) in hospitalized patients. Res Social Adm Pharm. 2025.
3. Systematic review of AI-based models in pharmacoepidemiology for adverse drug event prediction and detection. Front Drug Saf Regul. 2026;6:1773186 (PROSPERO CRD420251159394).
4. Chhikara P, Hammad TA. Rethinking drug safety signal detection and causality assessment in the age of AI: the risks of incomplete data and biased insights. Front Drug Saf Regul. 2025.
5. AI-driven pharmacovigilance: Enhancing adverse drug reaction detection with deep learning and NLP. J Med Surg Public Health. 2025.
6. AI-driven pharmacovigilance: Enhancing adverse drug reaction detection with deep learning and NLP (graph neural network framework for serious adverse event detection). PMC. 2025.
7. AI in Pharmacovigilance: Automating Adverse Event Detection. IntuitionLabs comprehensive overview, updated Feb 2026.
8. Adverse drug event detection using natural language processing: A scoping review of supervised learning methods. J Biomed Inform / PMC. 2023.
9. Weissenbacher D, Sarker A, Paul M, Gonzalez G. Overview of the Social Media Mining for Health (SMM4H) Shared Tasks at EMNLP 2018. Proc 2018 EMNLP Workshop SMM4H. 2018:13-16.
10. Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, McDermott M. Publicly Available Clinical BERT Embeddings. arXiv:1904.03323. 2019.
11. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36(4):1234-1240.
12. Real-world safety of tirofiban: a disproportionality analysis using data from FAERS and WHO-VigiAccess. PMC. 2025.

13. Characterization of ocular adverse events associated with crizotinib: real-world insights from FAERS and VigiBase. PMC. 2025.
14. P-1973. Signal Detection of Neurotoxicity Associated with Carbapenem Use: A FAERS-Based Disproportionality Analysis. Open Forum Infect Dis / IDWeek abstract. 2025.
15. Drug-induced glucose metabolism disorders: A disproportionality analysis based on the FAERS database. medRxiv. 2025.
16. Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. J Am Med Inform Assoc. 2020;27(1):39-46.
17. Disproportionality analysis of drug-associated progressive multifocal leukoencephalopathy using spontaneous reports: a 20-year signal detection study based on the FAERS database. PMC. 2025.
18. NADE: A Benchmark for Robust Adverse Drug Events Extraction in Face of Negations. arXiv:2109.10080. 2021.
19. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol. 1998;54(4):315-321.
20. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001;10(6):483-486.
21. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat. 1999;53(3):177-190.
22. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf. 2002;11(1):3-10.
23. An agency of the European Union: AI in pharmacovigilance - EMA update. 19th Industry Platform on EU Pharmacovigilance presentation. European Medicines Agency. 2025.
24. TransCelerate AI Pharmacovigilance: FDA and EMA Roadmap. IntuitionLabs. 2026.
25. Regulating the Use of AI in Drug Development: Legal Challenges and Compliance Strategies. Food and Drug Law Institute (FDLI) Update. 2025.
26. AI in Pharmacovigilance: 2025 Guidelines from EMA, FDA and ICH. Vitrana industry analysis. 2025.
27. EMA and FDA issue joint AI guidance for medicine development. European Pharmaceutical Review. 2026.
28. Benchmarking Local LLMs for Natural-Language-to-SQL Querying in Biopharmaceutical Manufacturing: An Empirical Benchmark on Consumer-Grade Hardware. arXiv:2606.01338. 2026.
29. Leveraging Transformer Models for Enhanced Pharmacovigilance: A Comparative Analysis of ADR Extraction from Biomedical and Social Media Texts. AI (Basel). 2025;6(2):31.
30. AI for Pharmacovigilance: Revolutionize 2025 Safety. Lifebit industry guide. 2026.
31. CONORM: Context-Aware Entity Normalization for Adverse Drug Event Detection. medRxiv. 2023.
32. Fine-Tuning Transformer Models for Adverse Drug Event Identification and Extraction in Biomedical Corpora: A Comparative Study. In: Proc. Int Conf Intelligent Systems and Computing. Springer; 2023.
33. Increasing Adverse Drug Events extraction robustness on social media: case study on negation and speculation. arXiv:2209.02812. 2022.
34. Portelli B, et al. Improving Adverse Drug Event Extraction with SpanBERT on Different Text Typologies. arXiv:2105.08882. 2021.
35. Ahire YS, Patil JH, Chordiya HN, Deore RA, Bairagi VA. Advanced Applications of Artificial Intelligence in Pharmacovigilance: Current Trends and Future Perspectives. J Pharm Res. 2024;23(1):23-33.

HOW TO CITE: Sakthikumar P, Integrating Artificial Intelligence and Natural Language Processing into Automated Adverse Event Signal Detection: A Comprehensive Review of Frameworks, Real-World Data, and Regulatory Challenges, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 8, 1492-1517. https://doi.org/10.5281/zenodo.21843487