Qualifying AI-Enabled Pharmacovigilance Service Providers

A Practical Due-Diligence and Oversight Framework

Reinhold J. Schilling
Head of Global Pharmacovigilance and EU QPPV

Schilling Vigilance White Paper Series | No. 8 | Revision 0 | 25 September 2026

Scope: EU human-medicines MAHs. This operational framework is not legal advice, an audit standard, a company SOP or a universal AI-validation protocol. Regulatory status assessed at 25 September 2026.

Executive Summary

A provider may have a mature quality system and still operate a safety pathway that silently loses a report. An AI classifier may reject a contact before it reaches a reviewer; a model update may alter a validated decision; an interface may fail without leaving an item in the PV queue. Conventional supplier qualification alone does not establish control of these interactions. The central thesis of this paper is: Qualify the whole safety-information pathway, not the vendor’s AI claim.

The marketing authorisation holder (MAH) may subcontract PV tasks but retains responsibility for the effectiveness of its PV quality system in relation to those tasks and for the completeness and accuracy of its pharmacovigilance system master file (PSMF). The QPPV must have effective access to the information needed for system oversight. A contractor has its own agreed responsibilities; neither its contract nor a human sign-off replaces MAH oversight. [1–3]

The author-proposed qualification unit is Provider × Service × Technology × Context × Risk × Controls × Evidence. Answer–Evidence–Risk–Action+ turns each provider claim into an evidenced MAH decision with an owner, deadline and fallback. EMA and CIOMS material supports risk-based, context-specific assessment of the complete human–AI workflow, not these particular author-proposed categories. [9–11]

Eight actions: (1) inventory AI actions and upstream dependencies; (2) approve each intended use; (3) find silent-negative routes; (4) test the production configuration on representative inputs and difficult cases; (5) map PV-task delegation separately from data-processing and technology chains; (6) contract change, access, incident, fallback and exit controls; (7) monitor denominated errors, transfers and reviewer corrections; (8) requalify on consequential change or failure.

The AI-Enabled PV Service Provider

A PV operations provider may extract, classify, code or submit ICSRs. A surveillance or medical-writing provider may screen literature, prioritise signals or draft reports. A front-door provider may operate medical information, contact centres, product-complaint intake or chatbots. A specialist may provide translation or analytics. A technology supplier may provide a PV database, automation platform or integration without performing an assigned PV task. A model or cloud company may be an indispensable upstream dependency. One organisation may occupy more than one category.

Distinguish delegation of a PV activity from technology provision, data access and personal-data processing. An affiliate or translator performing part of an assigned task may be a further PV-task subcontractor. A cloud host is not automatically one simply because it processes data. The GDPR processor and subprocessor analysis is separate, while operational assurance must include any fourth party that can change safety decisions, data integrity or service continuity. [2,6]

Regulatory Accountability Cannot Be Outsourced

Implementing Regulation 520/2012, read with its applicable 2025 amendment, addresses quality-system effectiveness, records, the PSMF, audits and outsourcing. Regulation 2025/1466 specifies subcontract terms covering allocation of roles, relevant safety-data exchange, and audit and inspection arrangements; further subcontracting of an assigned PV task requires the MAH's prior written consent. These are scoped legal provisions, not a reason to label every software vendor a PV-task subcontractor. [1,2]

GVP Module III Revision 2 has been effective since 10 September 2026 and describes inspection of direct and further PV subcontractors. It is guidance about authority inspection, not a universal prescription for the form of every MAH supplier audit. Other GVP modules should be read in light of the amended legislation and EMA's update information. QPPV access, retrievable records, a current PSMF and an effective quality system must work in operation, not only in contracts. [1–4]

Source	Status	Appropriate use
520/2012 as amended by 2025/1466	Binding PV law	Quality system, PSMF, subcontract terms, records and audits [1,2].
GDPR	Binding where applicable	Fact-specific roles, security and international transfers [6].
EU AI Act as amended	Binding; staggered	Classify actual use and operator roles; no automatic PV high-risk label [7,8].
GVP Module III and GVP index	Final guidance / index	Inspections and current module status [3,4].
ICH E2D(R1) EU strategy	Implementation guidance	Relevant source classification and transition [5].
EMA AI paper; EMA/FDA principles; CIOMS	Non-binding reflection / principles / consensus	Context-specific assurance, not a vendor checklist [9–11].
Revised Annex 11 / proposed Annex 22	GMP consultation drafts	Cross-domain analogy only, not adopted PV law [12].

The AI Act has different scope and commencement rules. Regulation 2026/1744 amended the original timetable; specified Annex III high-risk rules are scheduled for 2 December 2027 and specified Annex I product-related rules for 2 August 2028. These dates do not describe every AI Act provision. Classification requires a case-specific legal assessment. [7,8]

Start with Intended Use

"AI improves efficiency" says nothing about whether a tool drafts a field, suppresses an item, changes routing or closes an interaction. An intended-use record, proposed by the author, should identify the actual decision point and its consequences before a demonstration or performance claim is assessed.

Record field	What to specify
PV task and AI action	Find, extract, translate, draft, code, rank, exclude, transfer, submit or close.
Inputs and outputs	Original sources, eligible population, output recipient and regulated-record status.
Authority and negative route	Decision/release owner; false-negative consequence; visibility of rejected items.
People and operating envelope	Original-source access, reviewer capacity; products, countries, languages, hours, volumes and interfaces.
Configuration and chain	Provider, model/version, prompts, retrieval, thresholds, sites, support, telemetry, backup and subtiers.
Monitoring and recovery	Error sampling, reconciliation, manual route, rollback, record retention.

A provider that cannot identify which inputs may be excluded cannot supply a reliable denominator for a missed-item measure. Approval must attach to a defined service, configuration and context, not a generic tool licence. [9–11]

Qualification Lifecycle

The following author-proposed lifecycle combines related decisions; each stage needs an MAH owner, an evidence record and a disposition. Scale depth to the consequence and detectability of errors.

Stage	Evidence and owner	Response to a gap
Scope and inherent risk	PV owner/QPPV: task, autonomy, harm, silent negatives and dependency map.	Clarify or limit intended use.
Prequalification	Vendor management/Quality: entity, capability, QMS, AI disclosure and sites.	Hold if performer or AI action is concealed.
Questionnaire and records	PV/Quality/IT: answers linked to completed transactions, SOPs and system records.	Request evidence, not further assurances.
Demo and proof of concept	PV/IT: assessor-selected inputs, exact version, positive and rejected cases.	Retest or restrict if critical miss unbounded.
Assurance and audit decision	Quality/IT: validation reports, exceptions and risk-based audit rationale.	Audit before high-impact release where evidence cannot otherwise be verified.
Contract and transition	PV/Legal/Privacy: executed terms, source reconciliation, staffing and fallback drill.	No go-live with unresolved critical handoff.
Release and operation	PV/Quality: approved version, denominator-based monitoring and QPPV visibility.	Increase review or revert on adverse trend.
Change, incident and requalification	PV/Quality/IT: impact, affected population, CAPA and regression.	Manual operation, restrict or suspend.
Exit and support	PV/IT: tested export, open-work ledger, reconciliation and archive access.	Delay cutover or deletion.

Provider Claims Are Not Evidence

A marketing statement and questionnaire response are assertions. A policy states an intended practice; a contemporaneous record shows its execution. Version-matched testing assesses a defined configuration, while independent verification can challenge both method and result. The author-proposed Evidence Quality Ladder is not a regulatory grading system.

Level	Evidence
0	None
1	Assertion
2	Policy, SOP or vendor documentation
3	Completed contemporaneous operational record
4	Representative version-matched test or production evidence
5	Independent verification or audit evidence

A high nominal level is not necessarily relevant. A broad security certificate cannot establish detection of safety information; a recent production negative-route sample may be more useful. Check provenance, tested scope, version, recency, population and exceptions. A curated demo may use clean cases, hidden manual edits, a demo-only stack or positive examples alone. It does not establish validation or the safety of unseen rejections. [9–11]

Performance in the Intended PV Context

Assess the base model, configured application, routing workflow, combined human–AI process, delivered service and regulatory outcome separately. Success at an early layer does not prove that a case reached PV on time or that a reviewer intercepted an error. The joint EMA/FDA principles support examination of the complete human–AI system. [10]

Define the unit and eligible population, independent reference standard, adjudication process, separation of training, tuning and holdout data, production version and material subgroups before testing. Use both prevalence-representative work and enriched rare-event challenges; the latter reveals failure modes but cannot alone estimate routine precision. Include poor scans, mixed languages, ambiguous narratives and realistic workload. Report uncertainty, error severity and reviewer capacity. [9,11]

Recall or sensitivity equals correctly detected relevant items divided by all relevant items; precision equals correct detections divided by all detections; false-negative rate equals missed relevant items divided by all relevant items. Positive-only QC cannot estimate recall. A headline accuracy percentage can mask rare serious misses or weak language performance. No universal acceptance threshold is asserted. [9,11]

Answer–Evidence–Risk–Action+

The author-proposed Answer–Evidence–Risk–Action+ method records: Answer, the exact provider claim; Evidence, objective and version-relevant support; Risk, inherent and residual consequences after proven controls; Action, accept, clarify, remediate, contract, monitor, audit, restrict, reject, suspend or exit; Plus, an MAH owner, due date, verification, reassessment trigger and manual or rollback route.

Example: a provider says qualified staff review chatbot cases. The logs show that only AI-positive chats reach staff. The residual risk is an unseen safety-negative conversation. The MAH should hold autonomous negative closure, preserve all chats, test a blinded negative sample and route uncertainty to people; the PV owner verifies the fix before release and switches to manual routing if the control fails.

The method makes assumptions and decisions inspectable, but it cannot repair poor reference truth or replace clinical judgement. A critical gap in safety capture cannot be averaged away by excellent unrelated security scores.

Practical Due-Diligence Matrix

This author-proposed decision matrix samples the most consequential pathways. “Strong” means suited to this service and version; a red flag calls for the stated control decision, not a lower composite questionnaire score. Abbreviations: AE, adverse event; PV, pharmacovigilance; MAH, marketing authorisation holder.

MAH question	Acceptable answer	Strong evidence	Weak evidence	Yellow	Red	MAH action
AI disclosure: where used?	All actions mapped	Live inventory	Brochure	Minor gap	Hidden decision	Isolate; hold
Intended use: which authority?	Task and release owner named	Approved flow	Efficiency claim	Broad scope	Unknown authority	Define; restrict
Silent exclusions: which unseen?	Rejects logged and checked	Blind negatives	Positive demo	Narrow sample	Unlogged discard	Stop exclusion
Performance: what measured?	Metric and denominator defined	Adjudicated outcomes	Accuracy slogan	Aggregate only	No truth set	Require study
Representativeness: whose cases?	Relevant sources and languages	Dataset manifest	Curated inputs	Sparse cohort	Unrelated data	Limit or retest
False negatives: how found?	Independent negative review	Reject audit	Positive QC	Small sample	No detection route	Manual or shadow
Human review: what visible?	Originals, time, authority	Observed tasks	Human label	High workload	Summary only	Redesign
Model version: what deployed?	Version and updates recorded	Release manifest	Model name	Late register	Silent update	Pin or restrict
Prompt/retrieval: who edits?	Controlled diffs and tests	Release history	Prompt sheet	Manual log	Live free edits	Freeze and test
Reproducibility: trace a case?	Source to action retrievable	Case replay	Screenshot	Partial history	No reconstruction	Remediate/stop
Production monitoring: what alerts?	Segmented errors and drift	Investigated trends	Monitoring SOP	New baseline	No review	Enhanced QC
Customer training: allowed?	Explicit controlled position	Terms and settings	Verbal pledge	Telemetry unclear	Unapproved reuse	Block route
Upstream retention: what persists?	Logs and outputs governed	Provider terms	EU-hosting label	Backup unknown	Unknown reuse	Minimise/hold
Locations/access: where?	All processing/access mapped	Flows and logs	Region name	Support omitted	Unknown recipient	Privacy hold
Subprocessors: who else?	Current tiers disclosed	Inventory/terms	Top-tier list	Stale register	Hidden recipient	Map/control
PV subdelegation: consent?	Assigned tasks consented	Written approval	General notice	Consent pending	Unconsented task	Stop delegation
Audit trail: what survives?	Timed source/actions kept	Historic retrieval	SOP	Slow retrieval	Missing trail	Repair/reject
Manual fallback: staffed?	Tested with timestamps	Timed drill	Paper plan	No surge trial	No usable route	Hold release
Continuity: restored intact?	Queues and records reconcile	Restore exercise	Tabletop	Partial restore	Unreadable records	Repair/retest
Incident/CAPA: bounded?	Lookback and effectiveness	Event files	Ticket policy	Closure only	Serious miss ignored	Audit/suspend

MAH question	Acceptable answer	Strong evidence	Weak evidence	Yellow	Red	MAH action
Change: who authorises?	Risk-based approval/rollback	Release tests	Release notes	Narrow test	Unilateral change	Revert/restrict
Inspection: actual access?	Rights and retrieval workable	Executed terms/drill	Certificate	Room limits	Access refused	Reject route
Portability: what export?	Originals and metadata readable	Trial export	Exit promise	Costly format	Irretrievable trail	Negotiate/reject
Migration: complete?	Object-level exceptions resolved	Source-target match	Overall count	Attachments missed	No reconciliation	No cutover
Chatbot negatives: sampled?	Negative/abandoned saved	Blind transcript audit	Satisfaction score	Positives only	Chats disappear	Manual route
AE + complaint: both routed?	Two receipts where required	Seeded dual test	Process map	Late receipt	PV misses injury	Contain/retest
Requalification/correction load?	Triggers and burden reviewed	Review/time logs	Annual promise	Rising burden	Critical repeat	Resource/restrict

Legal, privacy and clinical subject-matter reviews remain necessary. The matrix is a decision aid, not a certification scheme or a substitute for examining sampled production records. [1,2,6,9–11]

Subcontractors and Fourth Parties

The author-proposed dependency map runs from MAH to direct PV provider, affiliate, call centre, freelancer, translator or annotation partner, then through application, cloud, foundation model, logging, security and remote support. Record each entity, task, access, site, upstream change rights, incident path, audit reach and exit alternative. The amended PV written-consent rule concerns further subcontracting of assigned PV tasks; GDPR onward processing has separate rules. A technology provider may be operationally critical without becoming a PV-task delegate. [2,6]

If the direct provider cannot obtain upstream model weights or proprietary test records, ask instead for identity and version, data-use terms, update/retirement commitments, independent black-box challenges, scoped third-party assurance, monitoring and substitution rights. If critical behaviour can change invisibly and no effective regression, monitoring or fallback is possible, do not release a high-impact exclusion route. [9–11]

Data Locations, Privacy and Confidentiality

An EU storage region does not answer where inference occurs or where staff may connect. Map primary storage, temporary processing, inference, remote support, prompt/output logs, telemetry, backups, disaster recovery and onward model-provider handling separately. Document retention and deletion for each copy.

Controller/processor status, lawful basis, special-category conditions, possible impact assessment and Chapter V transfer mechanisms depend on the facts and need legal review. Check provider and upstream settings before permitting identifiable PV content in an AI service. Customer-data training, model improvement and public or unapproved tools need an explicit controlled position; pseudonymisation alone does not settle the GDPR analysis. [6,9]

Model, Prompt and Workflow Change Control

This author-proposed taxonomy requires impact, evidence, regression scope, MAH approval where justified, enhanced monitoring and rollback for each change. It does not set a universal notification interval.

Change category	Potential PV effect	Proportionate evidence and decision
Model/provider, fine-tune or training data	Classification and error pattern change	Compare versions on rare positives and negatives; prior MAH approval for material decision effects; pin or roll back.
Prompts, retrieval and thresholds	Different evidence, output or rejection	Diff settings; replay critical and rejected cases; approve changed routing before release.
Rules, interface, API or dictionary	Lost transfers, miscodes, reviewer behaviour	End-to-end reconciliation, coding and UI tests; monitor first production period.
New product, language, country or source	Unrepresentative tests, new local rules	Scoped dataset and competence review; authorise expanded scope.
Hosting, support or subcontractor	Access, continuity or inspection change	Privacy/security review and PV-task consent where required.
Emergency patch/workaround	Unbounded or temporary deviation	Contain, document, perform minimum safe regression, notify by agreed route and test return.

EMA’s reflection paper allows consideration of incremental learning in PV with appropriate validation, documentation and monitoring; it does not permit untracked changes. [9]

Risk-Based Audit Strategy

Questionnaire-only review may suffice for a low-impact aid with complete source review and strong records. Remote evidence review fits a stable, transparent workflow; focused virtual audit addresses one technical issue. Full remote or on-site multidisciplinary audit is appropriate where silent failure, complex tiers or serious uncertainty warrant it. Follow-up audits test CAPA; for-cause audits investigate incident or recurring failure. These are proposed MAH choices, not authority inspection categories. [1,3]

Audit triggers include unobservable negatives, tested/production mismatch, unknown upstream changes, missing audit trails, repeated material errors, denied access, failed continuity or an incident whose population cannot be bounded. The test library is short: assessor-selected positives and rejects; source-to-output and timestamp reconstruction; production-version/prompt history; reviewer actions; change and CAPA records; actual data routes; subcontract terms; failed-transfer and fallback drill; migration rehearsal. Controlled viewing can protect patient data without sacrificing inspection access.

Contractual Control Architecture

The allocation below is an agenda for specialist legal and operational review, not model clauses. The amended PV subcontract terms apply within their legal scope; additional technical controls are selected by risk. [2]

Instrument	Control topics
Main agreement	Entities, scope, governance, continuity, audit cooperation, termination and schedule hierarchy.
PV agreement / SDEA	Task ownership, source clocks, safety exchange, QPPV access, urgent escalation, further PV delegation, records and autonomous-action limits.
Data-processing agreement	Roles and instructions, subprocessors, access and transfers, data use/training, assistance, retention and deletion.
Security schedule	Privileges, segregation, encryption, logs, incident support and tested restoration.
SLA	Defined clocks, denominators, acknowledgements, reconciliation, backlog and escalation.
Technical / validation schedule	Approved intended use, model/prompt/retrieval manifest, evidence, regression, releases and rollback.
Exit schedule	Readable exports, open items, legacy access, reconciliation, support and deletion after acceptance.
Operational instruction	Reviewer source access, negative sampling, adjudication, queues and handoffs.

KPIs, KRIs and SLAs

The following are author-proposed metric definitions, not acceptance limits. Publish counts and rates, sample size and how each eligible input entered the denominator. Segment by relevant product, territory, language, channel, source, severity and version. Escalate a serious miss even when an aggregate rate appears favourable.

Metric	Numerator / denominator	Segment	Limitation or gaming	Escalation
Safety transfer timeliness	Recognised safety items delivered within agreed interval / independently identified recognised items	Source/channel	Unrecognised items absent; reconcile intake	Late or lost item: manual handoff
ICSR timeliness	Due cases completed on time / all cases due	Seriousness/country	Queue-only denominator misses uncaptured cases	Assess late cases
Critical/material error	Independently confirmed errors / audited eligible outputs	Field/task	Biased samples hide defects	Lookback for critical error
False-negative rate	Adjudicated relevant items missed / all adjudicated relevant items	Language/version	Positive-only QC invalid	Restrict on serious miss
Missed literature	Relevant missed articles / adjudicated relevant search hits	Source/language	Search coverage separate	Restore and review
Chatbot handoff sensitivity	Correctly handed-off safety chats / all adjudicated safety chats	Language/abandoned	Lost chats distort denominator	Manual route
AE and complaint dual routing	Dual-route items with both receipts / all adjudicated dual-route items	Channel	Send not equal receipt	Contain unmatched
Failed transfer	Failed or unreconciled attempts / all expected attempts	Interface	Retries conceal initial failure	Manual transfer
Acknowledgement failure	Transfers without required receipt / all requiring receipt	Destination	Delivery not acceptance	Trace urgently
Reconciliation discrepancies	Unmatched source/destination objects / all eligible source objects	Channel/period	Totals can mask mismatched cases	Investigate each material gap
Backlog ageing	Open items beyond agreed age / all open eligible items	Queue/severity	Throughput hides oldest	Deploy capacity
Drift review	Flagged inputs / all assessed inputs; adjudicated errors / sampled flags	Version/source	Drift alert not harm proof	Targeted QC
Override/correction	Materially corrected outputs / all reviewed outputs	Reviewer/severity	High rate can mean good detection	Assess errors and workload
Human Correction Burden	Active correction minutes / all reviewed items; separately / corrected items	Task/reviewer	Time estimates biased	Capacity and redesign
Incident notice	Complete timely notices / all trigger events	Event class	Hidden events excluded	Audit event detection
CAPA effectiveness	CAPAs passing effectiveness check / all CAPAs due	Cause	Closure not effectiveness	For-cause review
Change notification	Approved/notified changes / all actual material changes	Change type	Silent upstream change hidden	Release reconciliation
Migration reconciliation	Verified target objects / all source-manifest objects	Object type	Counts hide wrong content	No cutover/deletion

Human Correction Burden is an author-proposed oversight measure, not a regulatory metric. Interpret it with error severity, final outcome and staffing: rising review minutes may erase a claimed efficiency gain and delay urgent work.

Medical Information and Contact Centres

Map calls, email, forms, chats, product complaints and patient-support programmes by language and operating hours. Preserve original content, earliest receipt evidence, transcription and translation where used. Provide urgent out-of-hours escalation, negative-classification and abandoned-session review, PV and Quality dual routing, bilateral acknowledgement and failed-transfer reconciliation. Do not assume every enquiry is a valid ICSR or every website an organised data collection system; apply the relevant legal and E2D(R1) source classification to the actual programme. [1,3,5]

Control	Acceptable evidence	Red flag	KPI	Audit test
Earliest receipt/original	Raw source to record with clock	Clock resets at handoff	Transfer time	Trace out-of-hours call
Negative or abandoned chat	Retained, independently sampled	Chats disappear	Handoff sensitivity	Seed ambiguous AE
AE plus complaint	PV and Quality receipts	Quality only	Dual-route completeness	Trace injury complaint
Translation/transcription	Original and medical-meaning QC	Unreviewed text replaces original	Material errors	Test negation and seriousness
Failure and hours	Alerted queue, staff and fallback drill	No urgent receiving route	Unmatched age	Simulate failed transfer
Privacy	Access/retention/model terms and logs	Unapproved public tool	Data-use incident	Follow recording to deletion

Incidents and Suspension

Classify missed or delayed cases, serious-case misses, wrong classifications, missed literature, failed AE/complaint routing, hallucinated content entering a PV record, trail loss, unapproved model/prompt changes, drift, failed interfaces, unauthorised training, cyber events and outages of the service or fallback. For a material event, contain the route and preserve originals, version manifests, logs and first-known timestamps. Bound the affected population since the last known-good state, review or reprocess it, assess regulatory implications, record deviation/CAPA and effectiveness, then retest before return. If the affected population or urgent transfer cannot be bounded, use a capable manual route and consider suspension. [1,3,9]

Statutory deadlines are not generic supplier notice periods. Applicable ICSR clocks depend on the case and law. Under GDPR, a processor notifies its controller without undue delay after becoming aware of a personal-data breach; a controller reports to an authority when required without undue delay and, where feasible, within 72 hours. Agree separate risk-based PV incident notification arrangements so the MAH can act in time. [1,6]

Exit Readiness Before Signature

Exit is a qualification gate because unfinished cases and source evidence must survive the provider. Before signing, define a trial export of originals, attachments, timestamps, audit trails, open cases and follow-up, backlogs, signal work, validation and performance history, relevant prompts, configurations, decision rules, model-version records and interface details. Include structured case formats such as E2B(R3) where relevant, but do not assume the format carries every attachment or historic trail. [1,3]

Proprietary embeddings, retrieval stores and workflow logic may not transfer in executable form. Establish how a successor would reproduce the essential route and reconstruct historical decisions. Freeze a source manifest, match objects and exceptions to the target, preserve legacy read and inspection access, and authorise downstream deletion only after accepted transfer and applicable retention assessment. A PDF-only export without originals, metadata or open-item status is not an executable exit.

Periodic and Event-Driven Requalification

Reassess on a risk-based schedule and on new AI use, model/provider, language, country, site or data location; a subcontractor change; acquisition or restructuring; critical incident; repeated KPI or CAPA failure; inspection finding; breach; unexplained drift; failed recovery; system migration; or inability to supply evidence. Outcomes are continue, continue with actions, enhance oversight, focus reassessment, audit, restrict, use manual processing, suspend or exit. A calendar review does not replace a change-triggered decision. [1,3,9]

Eight Practical Scenarios

These are illustrative author-proposed decisions, not findings about actual providers. Each decision depends on the evidence available for that exact deployment.

Use	Claim / missing question	Evidence	Red flag	MAH decision	Oversight
ICSR extraction	“Accurate extraction”; are serious details lost in poor scans?	Source-linked field tests on degraded inputs	Silent serious omission	Assisted entry only pending test	Sample critical fields
Seriousness classification	“Humans check positives”; who checks negatives?	Blind negatives and retained decisions	Auto-closed negative	No auto-exclusion; lookback	Review serious misses
MedDRA coding	“Validated globally”; same dictionary/language?	Version-specific verbatim-to-code tests	Obsolete unrelated test	Restrict to tested scope	Targeted coder QC
Literature exclusion	“AI removes irrelevant papers”; can missed hits be found?	Retained search corpus and reject sampling	Discarded articles	Audit before autonomous exclusion	Monitor missed hits
Signal prioritisation	“AI ranks”; who reviews the bottom?	Lower-tail review and source reconciliation	Suppressed unexamined items	Advisory use only	Review lower-tail ageing
Aggregate drafting	“Expert signs”; are claims and omissions checked?	Source-linked claim/omission review	Invented safety statement	Drafting only; expert review	Sample final claims
Chatbot routing	“Auto-routes AE and complaint”; what if both?	Dual receipts and abandoned-chat tests	PV misses injury complaint	Hold route until retested	Reconcile all channels
AI translation	“Linguists review”; where does source text go?	Approved API, originals, bilingual challenge	Unapproved training or lost negation	Hold identifiable content	Monitor material corrections

Qualification Decision Model

The categories and critical gates below are author-proposed governance tools, not regulator-assigned grades. Every approval identifies provider, service, version and intended-use scope, evidence, residual risk, approvers and reassessment triggers.

Category	Meaning and governance
Qualified	Gates met; release defined use and monitor.
Qualified with conditions	Bounded non-critical gap; interim control, owner, due date and expiry.
Qualified for restricted scope	Only tested languages, products, channels or advisory actions.
Qualification pending remediation	No affected production use until evidence or CAPA accepted.
Audit required before qualification	No high-impact release until unresolved material assertion is tested.
Not qualified	Necessary gate unattainable; do not use proposed route.
Qualification suspended	Previously approved route stopped; fallback and retrospective review.
Exit initiated	Govern continuity, migration, archive and controlled termination.

The proposed critical gates are: lawful controllable data use; clear PV roles and required consents; reliable safety capture; source-to-decision traceability; proportionate fitness evidence; workable fallback; audit/inspection access; controlled change; bounded incident impact; and feasible exit. A critical deficiency cannot be averaged away by high scores in unrelated domains.

Inspection and Audit Readiness Checklist

Select records independently and seek demonstration of operation, not merely policies. Approximately 25 high-value tests follow:

1. Who performs each assigned task, including affiliates?
2. Which channels, countries and languages feed the service?
3. Where does AI recommend, exclude, route or close?
4. Which further PV tasks have MAH written consent?
5. Do signed terms match actual data exchanges?
6. Can the QPPV see significant metrics and incidents?
7. Does the PSMF reflect relevant outsourced arrangements?
8. Can reviewers demonstrate task/language competence?
9. What is the earliest source receipt time for a selected item?
10. Can its original be traced to final disposition?
11. Who samples negative and abandoned items?
12. Were serious and rare cases represented in testing?
13. Does production match the approved model/prompt manifest?
14. Can a historical output and human decision be reconstructed?
15. Can reviewers access the original rather than an AI summary?
16. Are overrides and critical corrections investigated?
17. What reconciles source messages and receiving receipts?
18. Can one AE/complaint be shown in both teams?

19. What happens to a failed or out-of-hours transfer?

20. Can one incident’s full affected population be bounded?

21. Did a selected CAPA pass an effectiveness check?

22. Which upstream entities can access prompts, logs or recordings?

23. Has manual fallback operated under relevant load?

24. Can an export preserve originals, attachments and trails?

25. Will open work reconcile and historic records remain available at exit?

Conclusion

Outsourcing can bring expertise and scale; AI can improve selected PV workflows. Neither supplier sophistication, a contract nor a nominal reviewer proves that the pathway is controlled. The MAH should qualify and oversee the complete route from original safety information to its final disposition, including unseen negatives, upstream changes and workable exit. Evidence, rather than confidence, determines whether to accept, restrict, remediate, suspend or leave the service.

References

1. European Commission. Commission Implementing Regulation (EU) No 520/2012 of 19 June 2012 on the performance of pharmacovigilance activities. Binding; read as amended. Official source.

2. European Commission. Commission Implementing Regulation (EU) 2025/1466 of 22 July 2025 amending Implementing Regulation (EU) No 520/2012. Binding; generally applicable from 12 February 2026, subject to stated exceptions. Official source.

3. European Medicines Agency and Heads of Medicines Agencies. Guideline on good pharmacovigilance practices (GVP), Module III: Pharmacovigilance inspections (Rev. 2). 4 September 2026; effective 10 September 2026. Final regulatory guidance. Official source.

4. European Medicines Agency. Good pharmacovigilance practices (GVP). Module index and update information, September 2026. Regulatory guidance index. Official source.

5. European Medicines Agency. EU implementation strategy of ICH E2D(R1) Guideline: Post-approval safety data: Definitions and standards for management and reporting of individual case safety reports. 26 February 2026. Implementation guidance. Official source.

6. European Parliament and Council. Regulation (EU) 2016/679 of 27 April 2016 (General Data Protection Regulation). Binding, as applicable. Official source.

7. European Parliament and Council. Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Binding as amended, with staggered application. Official source.

8. European Parliament and Council. Regulation (EU) 2026/1744 of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence. Binding amendment. Official source.

9. European Medicines Agency, CHMP and CVMP. Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. Adopted September 2024. Non-binding scientific reflection. Official source.

10. European Medicines Agency and US Food and Drug Administration. Guiding principles of good AI practice in drug development. January 2026. Non-binding joint principles. Official source.

11. Council for International Organizations of Medical Sciences. Artificial intelligence in pharmacovigilance: Report of the CIOMS Working Group XIV. Geneva, 2025; published 4 December 2025. Non-binding consensus report. Official source.

12. European Commission, Directorate-General for Health and Food Safety. Stakeholders’ Consultation on EudraLex Volume 4: Good Manufacturing Practice Guidelines: Chapter 4, Annex 11 and New Annex 22. 2025 closed consultation on draft GMP texts; not final PV-specific requirements. Official source.