

Trust, Not Accuracy:
Why Pharmacovigilance Teams Still Double-Check AI Output They Know to be Valid

Despite strong and improving accuracy scores, AI-powered adverse event processing is often still being routinely double-checked by reviewers, well beyond the expectations of quality control. Could the answer involve a structured way of inferring appropriate reliance from operational evidence, such as through a trust coefficient reflecting the specific context of AI use?

AI output being correct doesn't automatically equate to it being trusted. Some pharma companies that have been drawing on AI-powered adverse event processing for a while now are reporting a frustrating disconnect between the demonstrated accuracy scores and the way PV teams are responding to them. Too often, reviewers are found to be reopening cases already signed off by the system, when there is no logical reason to do so. Audits confirm that this is happening, even though there has been nothing formal from a quality control perspective to require it. So, what's going on and what can be done about it - to ensure companies derive the expected benefits from the technology's efficiency? Accuracy isn't the whole story.

To a large extent, reviewer behaviour hasn't caught up with the growing operational evidence around AI use in PV: although platforms are shown to be reliable, user behaviour suggests they can't quite make the leap to leaning on it.

There is a perception issue linked to the new technology's approach, too. Anyone accustomed to using rules-based systems will have come to expect the same input to consistently yield the same output. AI challenges that expectation by design. Yet, consistency has been set as the form and a model is seen to only get things 'right,' 19 times out of 20, it looks unreliable. Bolting on more rules to fix this just makes matters worse. This is because each new constraint shrinks the pool of cases the system can handle safely on its own. There needs to be a shift in perception, then, to align with what AI does and how it works. In other words, trust needs to be redefined or reframed, to ensure appropriate user confidence in AI-driven output.

Consider what often happens with AI-enabled case narrative writing, for example. Here, most of what reviewers go on to change (when they decide to check and edit AI's output) comes down to house style rather than clinical substance or factual accuracy. If each stylistic tweak is recorded as a correction, the numbers make it look like the model is underperforming. This distorts the real performance picture and feeds people's instinct to over-check everything.

Building a Structured Measure of Reliance

Retaining indiscriminate quality control over an AI-driven process defeats the point of process transformation, by restricting the realised benefits – such as freeing up the precious time of safety scientists. That underlying problem is one of trust, manifested through continued old behaviour. However confident teams claim to be in AI-assisted systems, this counts for little if they keep checking every single thing a system produces when there is demonstrably no need to.

A formal trust coefficient is one way to gauge this impartially. Instead of scoring AI performance in isolation, a trust coefficient could combine several operational inputs into one structured view: model performance, reviewer behaviour, the maturity of validation and governance, task risk and how ready the organisation is to lean on AI output more broadly. The idea isn't to produce a fixed calculation, but rather a repeatable answer around how much independent checking a validated system genuinely still needs in the given context. As the evidence for reliable performance grows, the measures would allow oversight to ease off, without any compromise to governance or risk.

The measure would stand apart from the operational KPIs already in use (which track things like performance, efficiency or compliance), quantifying how much reliance is justified in a validated AI system under a defined set of conditions. Design conversations already treat trust as a core principle for AI platforms, but nobody has yet turned it into an operational number that regulators, vendors and pharmaceutical companies could apply consistently. That, more than raw accuracy, is where the real work in AI operationalisation needs to happen over the next couple of years.

Greater Autonomy for AI, No Change in Accountability

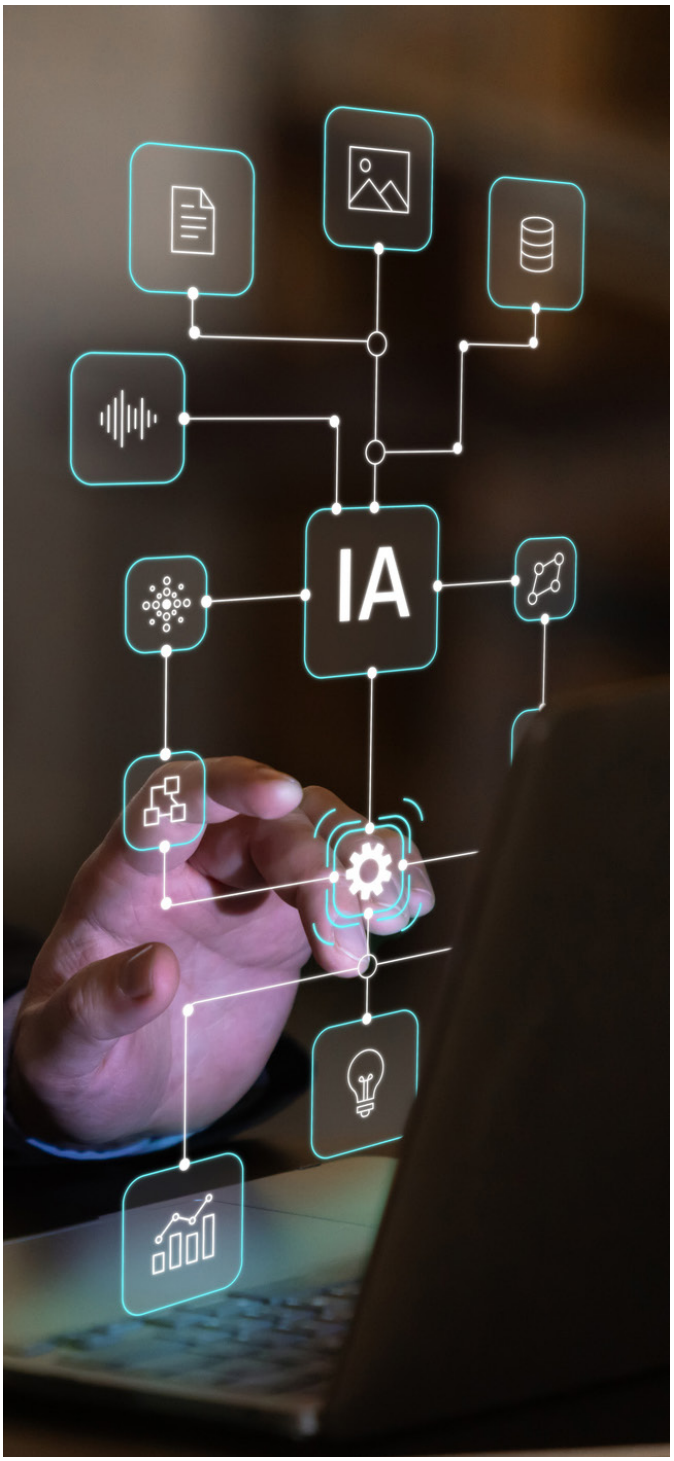
Agentic AI has a further bearing on any trust coefficient. Where earlier AI tools took care of single tasks in isolation, an agentic system can now run through a whole chain of preparatory work on its own, gathering data, checking fields, pulling together evidence, all before a human reviewer ever opens the case. This paves the way for specialists' attention to be concentrated on interpretation, judgement and the cases that are genuinely ambiguous enough to need a scientist's expertise.

The important point to highlight, amid that shift, is the difference between a reviewer checking every single step taken by a system and a reviewer stepping in only when risk or clinical significance actually warrants it. This doesn't alter ownership and accountability, however. Some organisations already have one AI model reviewing another's work before either output reaches a human being and none of that changes who is ultimately accountable. Agentic AI may change where the groundwork is carried out, but accountability for patient safety is not handed off with it; that remains the responsibility of qualified PV professionals.

Redesigning the Workflow Before Automating It

There is a wider point to take on board when building evidence-based reliance, for any organisation bringing AI into pharmacovigilance. And that is to redesign the process first and only automate once what's left is worth automating.

This is about ensuring that advanced automation solutions build on a set-up that has already been simplified, rather than try to retrofit



AI-powered capabilities to a process built for manual work. The ideal is to streamline hand-offs and touchpoints, but also leave room for future adaptation on the basis that AI capability itself is still evolving.

Regulatory Thinking Reflects the Need for Risk-Appropriate Intervention

Regulatory guidance is starting to reflect the same logic around AI risk and governance. Several regulators and international expert bodies have published new associated guidance this year, all pointing the same way: AI oversight needs to scale with risk and needs to be designed into a system from the start rather than bolted on once it is already running.

In January 2026 the European Medicines Agency and the US Food and Drug Administration published a joint set of 10 guiding principles for AI use across the medicines lifecycle, spanning early-stage evidence generation, clinical development and post-marketing safety monitoring.[1] The emphasis throughout is on validating systems in line with how they will actually be used, keeping data governance robust, watching performance on an ongoing basis and being open about where a system falls short. CIOMS took a comparable line in its own report on AI in pharmacovigilance that December, favouring governance built to flex as AI capability develops rather than rules pinned to today's specific technologies.[2]

As medicines become more complex and adverse event volumes continue to climb, the case for actively transforming ways of working in PV grows even stronger. VigiBase, the WHO's global repository of individual case safety reports, had logged upwards of 40 million entries by the close of 2024, roughly 70% of them added within the ten years prior.[3] That aside, volume itself is not the obstacle now. The outstanding challenge is determining when a validated system can be used with lighter checks.

Devising the optimal method for calculating a trust coefficient has yet to happen, but the justification for moving towards such a measure is strong, certainly for companies with substantial portfolios and big ambitions, which may be handling several hundred thousand safety reports a year.

REFERENCES

1. European Medicines Agency and U.S. Food and Drug Administration. EMA and FDA set common principles for AI in medicine development. https://www.ema.europa.eu/en/news/ema-fda-set-common-principles-ai-medicine-development-0, visited on 21 Jul 2026.
2. Council for International Organizations of Medical Sciences (CIOMS). Artificial Intelligence in Pharmacovigilance, CIOMS Working Group XIV report, Geneva. https://cioms.ch/working_groups/working-group-xiv-artificial-intelligence-in-pharmacovigilance/, visited on 21 Jul 2026.
3. Brand, J.S., Gauffin, O., Sartori, D., Fusaroli, M., Sköld, H., Bergvall, T., Sandberg, L., Wallberg, M., Hjelmström, P., Norén, G.N. VigiBase: Resource Profile Update with a Summary of Global Patterns and Trends in Adverse Event Reports for Medicines and Vaccines. Drug Safety, 49(6), 613–629 (2026).

Jason Bryant



Jason Bryant is General Manager of AI Platforms at ArisGlobal, bringing an AI-driven, human-centric approach to product design across pharmacovigilance. Trained originally as a data science actuary, his career has moved between fintech and health-tech, including a spell running a digital incubator at AstraZeneca before he turned his focus to agentic AI in drug safety. He is based in London, UK.

Email: jbryant@arisglobal.com.

Samuel Wallis



Samuel Wallis is Head of Case Processing at Bristol Myers Squibb, overseeing pharmacovigilance operations across one of the largest ICSR volumes in the industry. His transformation work at BMS has included moving the company's global safety database onto the cloud and rolling out AI-enabled case processing across the function. He is based in London, UK.