OMNITEK CONSULTING

AI Copilots in Medical Affairs

WHITE PAPER

Automating Literature Surveillance and MLR Triage Without Losing Scientific Rigor

Abstract

McKinsey estimates GenAI could generate $3 to $5 billion in annual efficiency gains for medical affairs functions, and the underlying capability is real: GPT-4 has demonstrated drug-information accuracy comparable to human pharmacists in peer-reviewed evaluation, and most organizations currently analyze less than 1% of their medical science liaison interaction data, leaving a substantial pool of clinical insight effectively untouched. At the same time, 2025 brought FDA enforcement letters at their highest volume in twenty-five years, and industry-wide, 95% of generative AI pilots are reportedly failing to deliver on their goals. The gap between the technology's genuine capability and its actual track record comes down to a specific, recurring mistake: treating AI-driven speed as the fix for MLR bottlenecks that are actually a governance problem, approved messaging scattered across shared drives, emails, and institutional memory rather than managed as a structured, verifiable asset. Generative AI trained against that ungoverned foundation introduces its own risk, semantic drift, subtle restatements that shift a claim's meaning without obviously contradicting it. This white paper applies established data-driven decision-support best practices to AI copilots in medical affairs, laying out how literature surveillance and MLR triage can be genuinely accelerated without the scientific rigor that makes medical affairs work trustworthy in the first place.

The Capability Is Real, and So Is the Failure Rate

Two things are simultaneously true about AI in medical affairs right now. The underlying capability has moved well past novelty: GPT-4 has shown drug-information accuracy comparable to human pharmacists in peer-reviewed testing, and FDA's own internal AI-assisted review pilots have reduced certain review tasks from three days to minutes. McKinsey projects $3 to $5 billion in annual efficiency value for medical affairs functions specifically from generative AI adoption.[i]

And yet organizational readiness lags the technology significantly. Only about 21% of organizations are actively evaluating GenAI use cases in this space, with another 19% in prototype stage and more than half still inactive. Industry-wide, a striking 95% of generative AI pilots are reportedly failing to deliver on their stated goals, a gap between capability and execution that deserves real scrutiny before any organization commits its medical affairs function to an AI-first strategy.

Why Faster MLR Review Does Not Fix What Is Actually Broken

The instinct behind most AI-driven MLR initiatives is straightforward: reviews are slow, so speed up the review. That instinct misdiagnoses the actual bottleneck. The real problem in most MLR processes is not reviewer speed. It is that approved messaging lives scattered across shared drives, emails, and institutional memory rather than as a governed, verifiable source of truth, which means content creators are drafting against sources they cannot confirm are current or authorized in the first place.

Generative AI applied on top of that ungoverned foundation does not fix it. It compounds it, through a specific failure mode worth naming directly: semantic drift, subtle restatements that shift a claim's underlying meaning without obviously contradicting the original approved language. A reviewer scanning quickly for obvious problems can miss drift that a careful line-by-line comparison would catch, precisely because the AI-generated version reads as plausible and well-formed.

The consequences of getting this wrong are not abstract. FDA issued more than 200 enforcement letters in 2025, the highest volume in twenty-five years, even as promotional material production rose 29% year over year. Adding to the irony, 77% of approved content is reportedly rarely or never actually used, evidence that the industry's content problem is not primarily a production speed problem at all.[ii]

Applying Best Practices: AI Copilots Built on Governed Content

A foundational best practice in data-driven decision support, surfacing genuinely actionable signal from noisy data, applies directly to medical affairs AI adoption. The organizations getting real value are not the ones deploying AI to review content faster. They are the ones treating approved messaging as governed, structured infrastructure first, then layering AI capability on top of that foundation.

1. **Approved Messaging as a Structured, Versioned Asset.** Treat every approved claim as a governed asset with clear provenance and version history, not a document buried in a shared drive, so any AI system referencing it works from a verifiable source of truth.

2. **AI-Assisted Literature Surveillance With Human Signal Triage.** Use AI to scan and summarize clinical literature at the volume no human team can match, while routing genuinely novel safety signals and unusual findings to human medical experts rather than treating AI summarization as the final word.
3. **Semantic Drift Detection as a Named MLR Control.** Build an explicit check for AI-generated content that compares meaning, not just keyword overlap, against the governed source claim, since standard reviewer scanning is not reliably designed to catch subtle meaning drift.
4. **MSL Interaction Data as an Underused Signal Source.** Apply AI-assisted analysis to the more than 99% of medical science liaison interaction data most organizations currently leave unanalyzed, surfacing emerging clinical themes at a scale manual review cannot reach.

None of this requires slowing AI adoption in medical affairs. It requires sequencing the work correctly: governed content and clear provenance first, AI capability layered on top of that foundation second, rather than expecting AI-driven speed alone to fix a bottleneck that was never really about speed.

What AI Copilots Buy, and Where It Gets Hard

Advantages

- **Insight capture at a scale manual review cannot reach:** AI-assisted analysis of the vast majority of MSL interaction data currently going unreviewed can surface clinical trends and unmet needs previously invisible to the organization.
- **Faster medical information response drafting:** response composition time can move from days to minutes for well-governed content categories, freeing medical affairs professionals for higher-judgment work.
- **Reduced review burden when governance is done first:** organizations that govern claims as structured assets before adding AI report meaningfully fewer content reviews and submission errors.
- **Broader literature surveillance coverage:** AI can scan and flag safety signals across a volume of published literature and social sources no human team could realistically monitor manually.

Risks and Barriers

- **Semantic drift is a genuine, hard-to-catch risk:** AI-generated restatements that subtly shift a claim's meaning can pass casual review while still constituting a real compliance exposure.
- **Ungoverned content makes AI adoption worse, not better:** layering AI onto scattered, unverifiable approved messaging compounds the underlying governance problem rather than solving it.
- **The 95% pilot failure rate is a real signal, not noise:** organizations rushing to deploy AI without addressing the governance foundation first are statistically likely to end up among the pilots that do not deliver.
- **Human-in-the-loop review cannot be treated as optional:** the high-stakes nature of medical affairs work means human oversight, source attribution, and escalation protocols for uncertain queries remain non-negotiable regardless of AI capability.

Case in Point: The Claim That Drifted Through Three Approved Versions

A mid-size specialty pharmaceutical company's medical affairs team adopted an AI tool to accelerate drafting of medical information responses to healthcare provider inquiries, generating first-draft responses referencing the company's approved clinical claims before human medical reviewers finalized them. Early results looked strong: drafting time dropped substantially, and the initial batch of AI-assisted responses passed review without incident.

Several months in, a routine audit of medical information responses compared a sample of AI-assisted drafts against the original approved source claims word for word rather than relying on the standard review process's faster scan. The audit found that one specific efficacy claim had drifted subtly across three successive AI-generated responses, each restatement individually plausible and each one passing standard review, but the cumulative effect had shifted the claim's practical meaning further from the original approved language than any single version would have suggested on its own.

No single reviewer had done anything wrong. The standard review process was built to catch obviously incorrect or off-label claims, not to detect a gradual, compounding meaning shift across multiple AI-generated iterations. The company's fix was not to abandon the AI drafting tool but to add a distinct semantic comparison step, checking each AI-generated response against the original governed claim language directly rather than against the most recent prior draft, closing the specific gap that had let drift accumulate unnoticed.

Recommendations for Medical Affairs and Compliance Leaders

1. **Govern approved messaging as a structured asset before deploying AI on top of it.** so any AI system references a verifiable, current source of truth rather than scattered, unmanaged documents.
2. **Build a dedicated semantic drift check into MLR review.** comparing AI-generated content against original approved claim language directly, not just against the most recent prior draft.
3. **Apply AI-assisted analysis to underused MSL interaction data.** to surface clinical insight from the vast majority of interactions currently going unanalyzed.
4. **Maintain human-in-the-loop review as a non-negotiable control.** for all AI-generated medical content, with clear escalation protocols for uncertain or novel findings.
5. **Treat the 95% pilot failure statistic as a governance signal.** auditing whether your organization's content foundation is structured enough to support AI before scaling a pilot further.
6. **Measure AI copilot success by insight and accuracy, not drafting speed alone.** since faster review that misses semantic drift is not actually a successful outcome.

Measuring Success: KPIs for AI Copilots in Medical Affairs

- **Approved messaging governance coverage:** the percentage of active claims managed as structured, versioned assets with clear provenance rather than unmanaged documents.
- **Semantic drift detection rate:** the number of meaning-level discrepancies caught by dedicated drift comparison versus standard review alone.

- **MSL interaction data analyzed:** the percentage of medical science liaison interaction data covered by AI-assisted analysis, tracked against the current sub-1% industry baseline.
- **Medical information response cycle time:** drafting and review time for medical information responses, tracked alongside accuracy to confirm speed gains are not coming at rigor's expense.

Conclusion

The capability behind AI in medical affairs is genuinely strong, and the value on the table, billions of dollars in efficiency gains and insight from data most organizations currently leave unanalyzed, is real. The 95% pilot failure rate industry-wide is not evidence the technology does not work. It is evidence that most organizations are deploying it on top of an ungoverned content foundation and expecting speed to solve a problem that was never really about speed.

Applying disciplined best practices to a governed-first approach to medical affairs AI, structured claims as the foundation, AI capability layered carefully on top, dedicated semantic drift detection, human oversight preserved as non-negotiable, turns AI copilots from a compliance risk into the genuine scientific rigor multiplier the technology is capable of being.

About OmniTek & Contact Information

OmniTek Consulting is a trusted partner for organizations seeking to enhance governance, operational agility, and decision-making. We specialize in modernizing business processes with a blend of human-centered design, technology expertise, and hands-on change management.

For medical affairs and compliance leaders evaluating AI copilots, OmniTek applies proven best practices to build a practical governed-first adoption practice: structured, versioned approved messaging, dedicated semantic drift detection, and AI-assisted insight capture from underused interaction data. Whether you are scoping your first medical affairs AI pilot or auditing one that has already launched, OmniTek can help you capture the efficiency gains without losing the scientific rigor your function depends on.

Contact us today to explore how OmniTek can support your transformation journey and drive lasting results.

McLean, VA | Serving clients nationwide
info@omnitekconsulting.com
omnitekconsulting.com

Click here to book a 15-minute call to see how OmniTek can help your team put these insights into practice.

Endnotes

[i] IntuitionLabs, "GenAI in Medical Affairs: Use Cases & Compliance Guardrails," citing McKinsey efficiency projections, MSL interaction data analysis rates, and GPT-4 drug-information accuracy research, available at https://intuitionlabs.ai/articles/genai-medical-affairs-compliance

[ii] pharmaphorum, "Is AI the Missing Link in Fixing MLR, or the Reason It Breaks?," on FDA enforcement letter volume, approved content utilization, and semantic drift risk in AI-generated promotional content, available at https://pharmaphorum.com/market-access/ai-missing-link-fixing-mlr-or-reason-it-breaks