

AI in Medical Affairs – An Independent Perspective

Where can LLMs improve the quality, efficiency, and strategic impact of Medical Affairs work most – and where do they still fall short?

Christian Wasmer · Consilience Strategy · March 10, 2026

www.consiliencestrategy.com

Introduction

The term “artificial intelligence” (AI) has become a catch-all for any software system designed to support or replace humans in executing cognitive tasks. In practice, however, AI encompasses a broad spectrum of technologies – from rule-based automation and predictive analytics to the generative language models (LLMs) that have captured public attention since 2023 (Fig. 1). This distinction matters as the capabilities, limitations, and appropriate use cases differ significantly across technologies. While this article will focus on the application and impact of LLMs in Medical Affairs and less on technological aspects, it is important to keep in mind the underlying probabilistic (non-deterministic) nature of LLMs. While not a requirement to understand this publication, the interested reader may find an introduction to LLM technology in section 2 of [1].

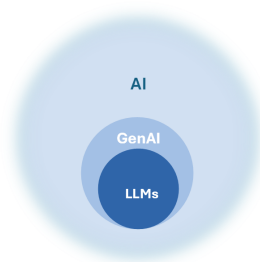


Fig. 1: LLMs are a specific form of Generative AI (GenAI), which in turn is a form of AI overall.

Technology companies developing AI tools have argued that LLMs, and eventually artificial general intelligence (AGI), will be capable of replacing humans at most knowledge-work tasks. While this view is obviously biased, today’s AI tools already demonstrate competence in certain domains: summarizing bodies of text, generating first drafts of structured documents, identifying patterns in complex datasets, and simulating conversational interactions. At the same time, they remain unreliable in others – particularly tasks requiring nuanced judgment, factual precision in specialized domains, or reasoning about causality. Whether a given task can be successfully carried out by an AI model depends on both the model’s specific properties (architecture, context length, training data, fine-tuning) and – more relevant for this discussion – the intrinsic characteristics of the task at hand (tolerance for error, need for

domain expertise, degree of ambiguity). In addition, the human level of familiarity and skill in interacting with LLMs introduces yet another variable (though that applies to any tool humans use).

The pharmaceutical industry in general, and Medical Affairs in particular, have been comparatively slow to adopt AI, LLMs and other digital tools, with several factors contributing: the highly regulated nature of pharmaceutical communications, the complexity of integrating AI into established workflows and review processes (e.g., medical-legal-regulatory, MLR), and legitimate concerns about accuracy and compliance in a domain where scientific rigor should always be more important than efficiency and speed. Yet organizations that cautiously and deliberately adopt AI workflows are likely to outperform competitors who either ignore these technologies or deploy them recklessly.

In this first part of a series, we examine LLMs' potential to directly support core Medical Affairs activities, from medical strategy and evidence planning to stakeholder engagement and publications. While specific applications will be addressed in more detail in future installments, our aim here is to answer the question: Where can LLMs most improve the quality, efficiency, and strategic impact of Medical Affairs work, and where do they (still) fall short?

To this purpose, we will group LLM capabilities into three functional categories:

Content Discovery

Search, retrieve, and extract insights from literature and data. This may include identification of the most relevant sources and sections within sources, but also creation of summaries, even across publications (i.e. some degree of content creation). For example, an LLM can rapidly scan thousands of congress abstracts to identify and summarize emerging data relevant to a very specific indication and/or asset – a task that would take days to complete manually.

Content Creation

The capacity to generate, draft, or adapt text-based outputs such as medical summaries, standard response documents, manuscripts and plain language summaries. Here, LLMs may act as accelerators, refining, expanding, and restructuring initial drafts with human input and guidance (“human-in-the-loop”).

Conversation Simulation

The use of LLMs to simulate realistic dialogue, enabling applications such as interactive internal training, training MSAs for HCP interactions/ objection-handling, or testing materials in simulated advisory board settings. This is arguably the most novel, least tested, but also potentially transformative category, as it aims at a dimension of Medical Affairs work – interpersonal communication – that has historically been relatively resistant to technological augmentation [2].

How LLMs are used at Pharma companies today

Most large pharmaceutical companies have started using LLMs in their workflows – though depth and sophistication vary. A 2025 analysis of the 50 largest pharma companies by market capitalization found that internal LLM platforms have become a standard pillar of enterprise AI strategy, alongside AI-integrated manufacturing facilities and drug discovery platforms. AI readiness has shifted from a differentiator to a baseline expectation [3].

In practice, deployment patterns cluster around a few recurring models. Companies like Pfizer have built internal LLM platforms (Pfizer’s Amazon-powered “Vox”) that serve as enterprise-wide research assistants for document retrieval and data querying. Merck and Bayer have deployed LLM systems across business units. Sanofi has positioned itself as an “AI-first” organization, claiming to embed AI across R&D, manufacturing, and commercial operations, and to have trained over 2,000 executives in data, AI, and strategy execution [3, 4]. These platforms are primarily used for Content Discovery tasks: literature search, competitive intelligence, internal knowledge retrieval, and document summarization.

Content Creation applications are less mature but growing. Pfizer, AstraZeneca, and Novartis have claimed to be using GPT-4 (via Microsoft Azure) to draft clinical study reports, protocol templates, and regulatory documents. Third-party vendors are entering this space with purpose-built tools, e.g., IQVIA with its NLP platform for extracting insights from clinical notes and publications. For Medical Affairs specifically, the most common applications remain internal: drafting training materials, synthesizing field medical insights, and generating first-pass summaries of competitive data.

Conversation Simulation is the least adopted category. While the concept of AI-simulated HCP dialogues for MSL training has generated interest, documented deployments remain limited. The most visible early moves are adjacent: Daiichi Sankyo announced plans to integrate agentic AI for personalizing HCP and patient responses, and some organizations use LLM-powered chatbots for internal medical information queries. The gap between aspiration and implementation is widest here.

These use cases reflect how LLMs are used internally by Pharma companies. But AI may also reshape how evidence is disseminated. Patients, HCPs, and other external stakeholders may increasingly seek health information from LLMs directly, giving rise to “generative engine optimization” (GEO): structuring content so it surfaces in AI-generated responses. Scientific communication and education will increasingly be mediated by LLMs before reaching HCPs and patients, however questions around accountability and regulation remain unanswered to date. The situation adds urgency to a question Medical Affairs teams may want to answer: How can LLMs be leveraged, and where, specifically, can LLMs create the most value for Medical Affairs?

Value of LLMs by Medical Affairs activity

Medical Affairs activities typically range from strategy creation (narrative, medical plan, evidence plan, etc.), execution (field and congress engagement, publications, ad-boards, etc.) and cross-functional contribution and advice (value dossiers, clinical strategy, etc.).

Drawing on our experience in working with Medical Affairs teams, our hands-on experience with LLMs, and the intrinsic properties of LLM technology, we mapped these three LLM capability categories against 10 key Medical Affairs activities, providing a strategic assessment of where LLMs' potential value is highest.

For gauging AI impact, we considered two aspects of work: speed (doing the same work faster) and quality (producing a better output). Most current LLM applications promise an increase in speed which may also translate to efficiency. However, we believe that improvements in quality are also possible when using the latest generation of LLMs and may be more relevant for Medical Affairs. Activities where AI is likely to have an impact on speed and/or quality while maintaining existing workflows were classified as “evolutionary”. To qualify as a “step change” in the matrix below, an LLM application must enable/ inspire a workflow or outcome that was previously impractical or impossible, not merely improve an existing one.

To illustrate how we distinguish “evolutionary” from “step change,” consider two examples from the matrix. In evidence planning, an LLM can identify relevant studies and literature in a fraction of the time a human analyst requires. The same technology, provided with structured guidance (e.g., a specialized “skill”), may produce an evidence gap analysis, mapping published data against a target product profile and identifying unaddressed clinical questions with specialist human oversight. This analysis can also be refreshed almost instantly when new data emerge, enabling more dynamic evidence planning. The workflow remains largely unchanged yet with clear efficiency/ speed (and potential quality) improvements – hence we rated LLM content discovery as ‘evolutionary’ for Evidence Planning. In comparison, Conversation Simulation is rated as a ‘step change’ for HCP Engagement because it offers a dimension of preparation – practicing against a simulated physician with specific objections, therapeutic focus, and communication style – that no existing tool or process provides at scale.

As a general pattern, the potential impact of Content Discovery correlates with the volume of data an activity requires as input. Activities that depend on synthesizing large bodies of literature, competitive intelligence, or field insights stand to benefit most.

Overall, the matrix is intended as a framework for prioritizing experimentation and investment.

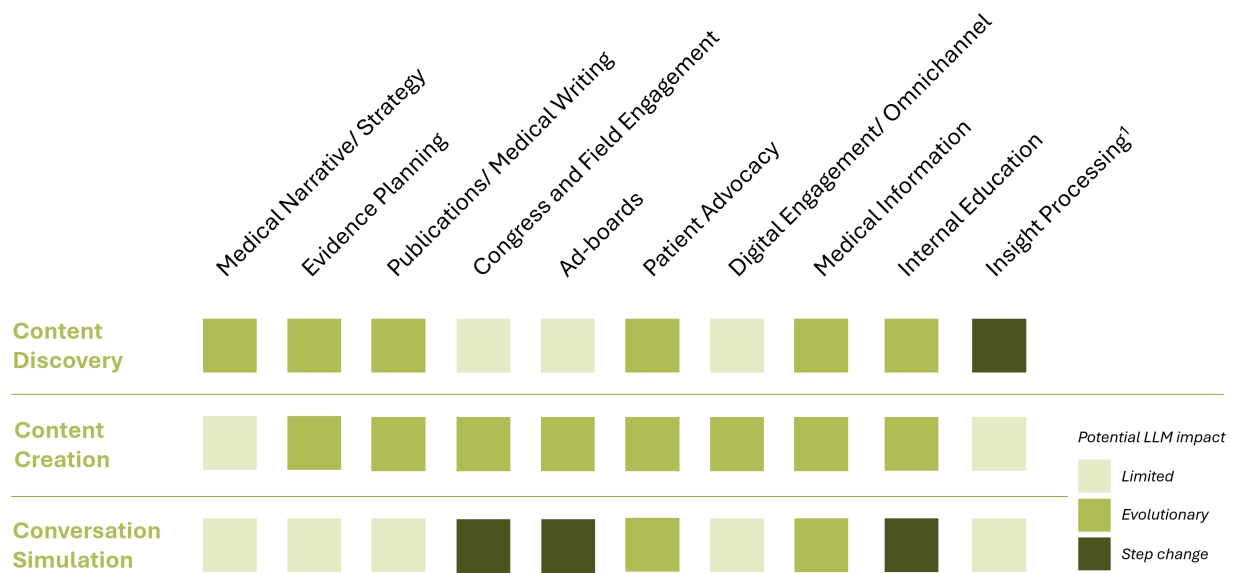


Fig. 2: Strategic assessment matrix evaluating AI’s potential value across medical affairs activities and three core AI capabilities (Content Discovery, Content Creation, and Conversation Simulation).

How to read this figure: For example, Content Discovery shows strongest potential AI impact (step change) for Insight Processing. Content creation capabilities have the potential to improve almost all medical activities. Conversation Simulation appears positioned for step-change impact specifically in (training for) HCP Engagement and Internal Education.

This framework is designed to guide implementation strategies by identifying where LLMs can incrementally improve existing workflows (evolution) versus where they might fundamentally transform how medical teams operate (leading to exponential and compounding gains).

Note: Some Medical Affairs activities were intentionally omitted from our assessment, e.g., Pharmacovigilance, Medical Affairs input for other functions’ activities etc., but those may still see improvements from LLM use.

¹ Insight Processing refers to the collective processing of insights collected from congresses, field engagement, patient engagement, digital engagement, and medical information in a compliant manner.

What Medical Affairs teams can do today

The previously discussed matrix reveals that not all potential LLM applications are created equal for Medical Affairs. It can guide where to start and what to pilot. And, as discussed, pharma companies are at different stages of the adoption process, so some examples will be in place already at some, but not all companies.

Content Discovery offers the broadest, most immediate value – it touches nearly every activity and carries relatively low compliance risk, since the LLM is retrieving and organizing existing information and, when prompted adequately, not generating novel claims (though without guaranteeing accuracy or completeness). If not in place already, this is where relatively inexperienced teams should start: piloting AI-assisted literature surveillance, congress coverage, and competitive intelligence. Unsurprisingly, these services are already offered by third-party vendors, though at varying levels of quality and

helpfulness. Summaries/ processing of sources may be using both discovery and limited content creation.

Content Creation represents the next tier of opportunity, but demands rigorous human oversight, particularly for externally used materials. Teams can begin with lower-risk internal applications, such as drafting training materials and internal education or synthesizing field medical insights, before scaling toward publications support. Counterintuitively, this is the category where LLMs may initially lower efficiency, e.g., if human review efforts increase faster than creation efforts decrease. While LLMs can learn and human-AI friction will decrease over time, managing expectations, providing ongoing guidance, and leveraging existing LLM experience within the team(s) involved will be critical for success.

Conversation Simulation is the category with the highest potential for step-change impact in HCP engagement and internal training. However, it is also the least mature and requires careful validation before deployment. A practical starting point: use AI-simulated HCP dialogues as a supplementary training tool for field teams, not as a replacement for real-world/ internal coaching. Another low-risk pilot of conversation simulation may be run for internal training, alongside the more traditional formats.

Teams who start experimenting across all 3 categories now – while maintaining clear governance and human-in-the-loop review – will be best positioned to reap the full benefits of LLMs as capabilities mature.

A word of caution on user satisfaction as a success metric: many LLMs are optimized – by design – to produce outputs that users find agreeable. This means that a team’s enthusiasm for an AI tool does not, by itself, indicate that the tool is improving the quality or efficiency of their work. An LLM that generates polished-sounding but substantively shallow medical narratives may receive high satisfaction scores from users while contributing nothing – or worse, introducing subtle errors – to the final deliverable. Organizations should therefore complement user feedback with outcome-based measures: Did the tool reduce time-to-completion for a defined task? Did it change the quality of the output as judged by an independent reviewer?

Getting Started: A Practical Guide

The preceding discussion identifies where value lies and what pitfalls to avoid. Translating this into action requires a structured approach and depends on a pharma company’s existing LLM capabilities and organizational context. Nevertheless, in the following we aim to provide a generic framework for moving LLMs from assessment to implementation:

- 1 Know your current AI infrastructure.** Understand what models, platforms, and capabilities are available at your organization today and in the near term – including whether custom applications or workflows can be built on existing systems.

- 2 **Select 2–3 use cases for structured pilots.** Use the matrix above to identify areas where AI capability aligns with Medical Affairs need. Prioritize use cases that are low-risk and high-frequency – e.g., literature surveillance, first-draft standard response letters, or competitive intelligence summaries – to build organizational confidence before tackling higher-stakes applications.
 - 3 **Train pilot teams and define success criteria upfront.** Provide specific guidance on what the AI tool can and cannot do, set expectations on quality review requirements, and agree on measurable outcomes (time saved, error rates, output quality as assessed by an independent reviewer) before the pilot begins.
 - 4 **Monitor through structured check-ins.** Set up weekly or bi-weekly touchpoints with pilot teams to capture what is working, what is not, and where workflows need adjustment. Pay attention to how people are using the tool, not just whether they like it (see above: user satisfaction is an unreliable proxy for impact).
 - 5 **Evaluate and decide at defined intervals.** Assess pilot performance against pre-defined criteria on a bi-monthly basis. Discontinue or redirect pilots that are not delivering measurable improvements. The goal is to learn (and sometimes fail) fast, not to justify a technology investment.
 - 6 **Scale what works; raise ambition.** Once two or more pilots have demonstrated measurable impact, expand those use cases across other Medical Affairs teams and begin piloting more ambitious applications – including those in the Content Creation and Conversation Simulation categories.
-

Beyond the Hype: Risks, Barriers, and Trends

This perspective was developed in the first quarter of 2026. Given LLMs' speed of evolution, some or most of what we write here may be outdated in one or two years. Therefore, we will close with only a brief note on key risks, barriers and current developments.

Perhaps the most consequential medium-term risk for Medical Affairs has only recently been described in detail and called “cognitive surrender by Shaw and Nave (2026). Their experiment demonstrated that when participants had access to an AI assistant, they adopted its outputs with minimal scrutiny, overriding both intuition and deliberate reasoning. Accuracy rose when the AI was correct and fell when it erred – the behavioral signature of cognitive surrender [5].

Medical Affairs depends on specialized judgment, and errors introduced early may propagate through publications, regulatory submissions, and HCP interactions, extending the consequences of uncritical adoption far beyond the individual user.

The risk is not limited to hallucinations, which can be caught with standard oversight. The medium-term concern is subtler: a decline in the depth of reasoning that Medical Affairs professionals bring to their interactions – with HCPs, with internal stakeholders, and in written communications. If LLMs do the thinking and humans do the approving, the quality of the underlying judgment erodes over time. Organizations deploying LLMs in Medical Affairs should monitor not just output accuracy but the reasoning capability of the humans in the loop.

Related to cognitive surrender are well-known but still unresolved questions around accountability, which we won't discuss here.

The cognitive surrender risk is compounded by the most significant current trend: the emergence of “agentic” AI – systems that autonomously plan, reason, and execute multi-step tasks rather than responding to single prompts [7]. Where current tools function as assistants, agentic systems can be assigned a goal and will attempt to independently gather data, cross-reference sources, draft outputs, and (ideally) flag issues for human review. While agentic AI may drive quality and applicability, the additional layers of autonomous action make it harder for humans to maintain the scrutiny that prevents cognitive surrender.


Additional bottlenecks remain on the more technical side: data confidentiality considerations, integration with legacy systems (Veeva Vault, MLR review workflows) not designed for AI interoperability, and sparse regulatory guidance – e.g., the EU AI Act classifies medical AI as high-risk [6], but detailed and Medical Affairs-specific rules are still emerging.

Yet despite these risks, we believe that a transition to LLM workflows is inevitable for at least some Medical Affairs activities. However, this trend will likely not reduce the need for skilled Medical Affairs professionals – on the contrary. While Tech company leaders have claimed that AI will soon replace highly skilled white-collar workers, the transition may at least lead to an increased demand for professionals such as Medical Directors who can exercise the judgment that prevents cognitive surrender. And in the long run it remains to be seen if substantial aspects of Medical Affairs work will be taken over by LLMs (or subsequent technologies), given how much quality and depth matter in our field.

In addition, as some Pharma operations may accelerate significantly due to AI (e.g., R&D), the broader application of such technologies may lead to increased demand for tasks requiring accountability, especially in the context of human-human interactions. A similar, seemingly counterintuitive effect is observed in another area of medicine: While image recognition and diagnostic algorithms outperform radiologists in specific imaging tasks, availability of the technology has coincided with an increased demand for radiologists as imaging volume has grown rapidly [8].

We will set out to explore specific LLM applications and gauge the potential impact on how Medical Affairs can contribute value in an ‘AI-interfused’ world in future parts of this series.

References

- [1] Lu J, Choi K, Eremeev M, et al. “Large Language Models and Their Applications in Drug Discovery and Development: A Primer.” *Clin Transl Sci*. 2025;18(4):e70205.
- [2] Fröling E, Domrös-Zoungrana D, Rajaeen N, et al. “Artificial Intelligence in Medical Affairs: A New Paradigm with Novel Opportunities.” *Pharmaceut Med*. 2024;38(5):331–342.
- [3] CB Insights. “Pharma AI Readiness: How the 50 Largest Companies by Market Cap Stack Up.” July 2025.
- [4] HEC Paris Executive Education. “The Sanofi Transformation: How 2,000+ Executives Are Driving a Data-Driven Revolution.” 2025.
- [5] Shaw SD, Nave G. “Thinking—Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender.” *PsyArXiv*. January 11, 2026. 

- [6] European Parliament. *Artificial Intelligence Act*. Regulation (EU) 2024/1689, approved May 2024.
- [7] Lakhan SE. "The Agentic Era: Why Biopharma Must Embrace Artificial Intelligence That Acts, Not Just Informs." *Cureus*. 2025;17(5):e83390.
- [8] Mousa D. "AI Isn't Replacing Radiologists." *Understanding AI*. September 25, 2025.

About the Author



Christian Wasmer is a Managing Partner and Co-Founder at Consilience. He has advised global pharmaceutical and biotech companies on medical strategy design and execution, clinical planning, commercial and portfolio strategy for over 10 years and previously worked at Prescient Healthcare Group, Trinity Life Sciences, Syneos Health, and the Boston Consulting Group.

CW_AI_P1@consiliencestrategy.com



Consilience is a boutique Life Science consultancy linking together science, strategy, and organizational context.

Our therapeutic area experience extends across oncology, hematology, rare disease, neurology, cardiovascular, metabolic disorders, and more. Technologies we're familiar with include (multi-specific) antibodies, allogeneic and autologous cellular therapies, antibody-drug conjugates, peptide-receptor-radionuclide-therapy and -diagnostics, protein degraders, and gene therapies.

The companies we work with range from Biotechs planning for first commercial launch to global Pharma companies. We typically work with Medical Affairs, Commercial, NPP, and Clinical function leads focused on an asset, technology, therapeutic area pipeline, or portfolio.

At Consilience, we're designing our own workflows with an 'AI first' mindset.

A note on how we work with LLMs

Development of this publication leveraged LLM support. The author defined the framework, selected sources, drafted the outline, and developed the figures. Claude Opus 4.6 proposed sources, expanded the outline into prose, and formatted the final output according to Consilience's brand guidelines. The two iterated through multiple rounds of revision. Consilience team members performed a final review and approved all content. Claude accelerated research and drafting, but every substantive judgment—direction, framing, source credibility—remained human.