*Public Management and Governance Review – 2025*
*DOI: https://doi.org/10.60733/PMGR.2025.04*


From the Classroom

# It's not the AI,… it's the way you use it: Making LLMs work in the real world

Raphaela Mayer-Negm

## Abstract

Pharmaceutical deviation and corrective and preventive action (CAPA) processes are broken—plagued by rushed closured, "human error" cliché and data silos. Can large language models (LLMs) fix this, or will they just automate bad habits? Based on expert interviews, this article reimagines LLMs not as decision makers but as devil's advocates—tools that provoke better thinking, expose bias and deepen investigations. The proposed five-step roadmap demands more than tech: strong data governance, human oversight, iterative validation, small-scale pilots and a cultural shift toward learning. The EU's draft Annex 22 on Artificial Intelligence (AI), released in July 2025 for public consultation, adds timely pressure—demanding traceability and accountability but risking innovation paralysis if applied too rigidly. The takeaway: LLMs could be transformative, but only if quality leaders have the courage to use them as catalysts for change—not as digital box-tickers in the same old broken system.

## Key words

Large Language Models (LLMs), Pharmaceutical Quality Assurance, Corrective and preventive action (CAPA), CAPA Management, Deviation Handling, Artificial Intelligence (AI), AI in Good (x) Practices (GxP), Human-in-the-Loop (HITL), AI Validation, Regulatory Compliance, Cultural Change, Quality Risk Management

## Acknowledgements

I would like to thank all experts who generously shared their time, experience and reflections. Their openness in discussing the everyday pain points of deviation and CAPA management—and their willingness to explore the unfamiliar territory of LLMs—greatly enriched the depth and relevance of this work.

## Extra information

This article is an executive summary of a thesis written for the Executive MBA program Health Care Management, at the Executive Academy of the Vienna University of Economics and Business. This article and the underlying thesis were conducted independently and were not affiliated with or carried out in cooperation with the author's current employer. The work reflects the author's personal expertise and perspective in the field.

*Public Management and Governance Review – 2025*
*DOI: https://doi.org/10.60733/PMGR.2025.04*


## Introduction

In the pharmaceutical industry, managing deviations and corrective and preventive actions (CAPAs) is both a regulatory requirement and a core driver of quality improvement. Yet, despite decades of quality management initiatives, systemic issues persist: investigations often remain superficial, documentation is burdensome and CAPAs are inconsistently defined or executed. As advanced technologies enter the quality domain, one question surfaces with urgency: Can large language models (LLMs) help organizations do better—not just faster?

This article explores that question based on findings from a qualitative study involving quality assurance (QA) professionals in the pharmaceutical sector. Drawing on six expert interviews, I examine how LLMs—such as ChatGPT or Copilot—are currently being used, more importantly, how they should be implemented to realize their potential in CAPA and deviation management.

The study builds on and extends existing literature in pharmaceutical quality systems, such as Rodríguez-Pérez's [1] critique of superficial root cause analysis (RCA), the ISPE's AI Maturity Model [2] and regulatory expectations from the European Medicines Agency (EMA) [3] and the U.S Food and Drug Administration (FDA) [4] regarding AI traceability and validation. The core insight is this: LLMs are not just digital assistants, they can challenge cognitive bias, support systemic thinking and foster more meaningful investigations. But this only happens if technology is embedded in the right governance, culture and quality framework.

This article focuses on implementation: what practitioners can do now to integrate LLMs into pharmaceutical QA systems in a safe and value-adding way. While the study methodology is qualitative, further detail on research design and analysis steps is available in the Appendix.

## Persistent challenges in deviation and CAPA management

Qualitative interviews with industry experts confirm long-standing critiques of deviation and CAPA processes. Despite the availability of structured tools and quality frameworks, systemic weaknesses persist:

- **Time pressure and closure focus:** Investigations are often rushed to meet timelines, sacrificing depth for compliance. This prioritization aligns with FDA observations regarding recurring citations for inadequate CAPAs [5].
- **Inadequate root cause analysis:** Experts described root cause analysis (RCA) as frequently superficial, relying heavily on the generic attribution of "human error" rather than systemic diagnostics—a critique echoed by Rodríguez-Pérez [1] and Husman [6].
- **Fragmented and siloed systems:** Disconnected documentation systems and non-standardized terminologies impede consistent CAPA execution and hinder the potential for meaningful analytics or automation.
- **Lack of preventive thinking:** The industry's reactive mindset often results in corrective actions that fail to address broader organizational or procedural failures, reinforcing Snee's [7] view that systemic insight is missing.

While these challenges are well documented, the novelty of this study lies in examining how LLMs could be strategically integrated to support—not replace—critical thinking and quality decision-making.

*Public Management and Governance Review – 2025*
*DOI: https://doi.org/10.60733/PMGR.2025.04*


## LLMs as analytical sparring partners, not decision-makers

The literature often presents LLMs as transformative agents capable of accelerating decision-making or uncovering hidden trends [8], [9], [10]. However, findings from this study challenge that optimism. Experts agreed that current LLM applications in pharma remain support-focused, such as assisting in documentation or preliminary signal detection (trending of deviations). Crucially LLMs were not seen as fit for standalone RCA or CAPA determination.

Instead, experts interviewed emphasized LLMs' potential to function as *devil's advocate* or analytical sparring partners—provoking alternative explanations and uncovering biases that may be overlooked in politically or culturally constrained environments. This conceptual shift—from LLMs as decision engines to reflective tools—opens a new space in the discourse on AI in quality management. It positions LLMs not as replacements for human judgment but as catalysts for better thinking, collaborative learning and hypothesis testing.

## A roadmap for responsible LLM integration in CAPA and deviation management

Based on both empirical data and theoretical grounding, the following five-step roadmap outlines how organizations can safely and effectively integrate LLMs into quality systems.

### *1. Build a robust data and governance foundation*

Experts consistently cite data quality, standardization and system connectivity as prerequisites for any meaningful LLM application. Siloed, inconsistent or incomplete data compromises not only model performance but regulatory defensibility.

*Recommendation:*

Conduct comprehensive audits of their deviation and CAPA datasets, standardize documentation structures and integrate relevant enterprise quality management software. Data ownership, traceability and change control mechanisms must be clearly defined. This approach aligns with EMA [3] and FDA [4] expectations for audit-ready, validated AI systems.

### *2. Establish human-AI oversight with clear role boundaries*

LLMs should augment, not replace, QA expertise. Expert interviewees unanimously endorsed human oversight, especially for tasks involving root cause identification or CAPA strategy development.

*Recommendation:*

Implement hybrid oversight models that delineate which tasks LLMs may assist with (e.g., drafting investigation summaries) versus those requiring QA approval (e.g., final RCA decision). Training must focus not just on how to use LLMs but on how to critically engage with them—cultivating AI literacy, process knowledge and reflective questioning.

### *3. Develop dynamic validation and continuous monitoring protocols*

Static validation frameworks are insufficient for adaptive systems like LLMs. Experts advocated iterative validation, scenario testing and real-time monitoring.

*Public Management and Governance Review – 2025*
*DOI: https://doi.org/10.60733/PMGR.2025.04*


*Recommendation:*

Adopt a dynamic validation model encompassing pre-deployment simulations, stress testing and post-deployment audits. Maintain version control and detailed audit logs, aligning with GAMP 5 [11] and ALCOA+ principles [12]. Ensure explainability and transparency throughout the model lifecycle.

### *4. Start small: pilot in low-risk areas before scaling*

Phased implementation emerged as a common theme. Experts recommended piloting LLMs in low-risk tasks to build trust, iterate on governance and minimize exposure.

*Recommendation:*

Initiate small-scale pilots in areas like text generation for minor deviations. Use these pilots to refine prompt engineering, human-AI workflows and validation protocols. This strategy reflects a risk-based approach endorsed by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) quality guideline Q10 [13] and FDA guidance [14]

### *5. Support cultural and organizational readiness*

Perhaps the most significant barrier identified by experts was not technical, but cultural. Experts described environments focused more on checklist compliance than on system learning or preventive action.

*Recommendation:*

Foster a quality culture where AI tools are seen as collaborators in learning—not just automation shortcuts. Leadership must drive this shift, framing LLMs as tools for hypothesis testing, dialogue and organizational introspection. Training should emphasize both behavioral and technical fluency.

## Regulatory outlook: The implications of Annex 22 on AI Use in GxP

GxP is an umbrella term for good practices in various management and governance areas, such as Good Manufacturing Practices (GMP), Good Clinical Practices (GCP), and Good Laboratory Practices (GLP).

Following the completion of this study, the European Commission released a draft version of Annex 22 in July 2025 for public consultation. This annex, part of the EU Volume 4 Good Manufacturing Practices (GMP), outlines expectations for the use of artificial intelligence in pharmaceutical quality systems and manufacturing contexts. Its publication is a significant regulatory milestone, offering structured guidance on AI validation, risk management, data governance and human oversight.

While Annex 22 marks progress in aligning AI use with GMP principles, it also introduces new layers of complexity. From the perspective of a QA professional and LLM enthusiast, one might adopt a devil's advocate stance: the Annex, while well-intentioned, risks codifying a rigid, overly conservative compliance posture that may stifle iterative experimentations—especially for emerging tools like LLMs that thrive on prompt variability and evolving datasets. If

interpreted too restrictively, Annex 22 could paradoxically limit the very innovation it aims to regulate.

That said, its emphasis on human accountability, explainability and validation strongly echoes the findings of this study. The alignment underscores that responsible LLM integration hinges less on technology itself and more on the governance structures and cultural readiness surrounding it. Future work will need to explore how the evolving regulatory landscape, including Annex 22, interacts with organizational change, human factors and the practical dynamics of AI deployment in GxP environments.

## Conclusion

LLMs offer significant promise for improving deviation and CAPA processes in pharmaceutical manufacturing, but their impact depends less on technological sophistication and more on implementation context. Governance, culture, training and validation all shape whether these tools become enablers of meaningful quality improvement or merely another layer of digital bureaucracy. This article provides a roadmap based on real-world insights and validated by existing literature. For practitioners seeking to experiment with LLMs, the key is to start small but think big. LLM adoption should be viewed not as a technological upgrade, but as a cultural and procedural evolution toward proactive, reflective and system-oriented quality management.

## Author bio

Raphaela Mayer-Negm is Head of Quality Operations, Boehringer Ingelheim RCV GmbH & Co KG, Vienna.

ORCID: https://orcid.org/0009-0008-5267-8821

LinkedIn: www.linkedin.com/in/raphaela-mayer-negm-51a4a4171

## Appendix

This article is based on a qualitative study conducted between March and May, 2025. In addition to an elaborate literature review, six in-depth semi-structured interviews were conducted with senior quality professionals across five pharmaceutical organizations. Participants were selected for their practical experience with deviation and CAPA management and their exposure to AI or digital tools in GxP settings. Interviews were transcribed and coded using inductive thematic analysis. The study prioritized depth over breadth, aiming to surface rich, contextualized insights into real-world use cases, constraints and perceptions.

## References

[1] J. Rodríguez-Pérez, *Handbook of investigation and effective CAPA systems*, Third edition. Milwaukee, WI: Quality Press, 2022.
[2] N. Erdmann, R. Blumenthal, I. Baumann, and M. Kaufmann, "AI Maturity Model for GxP Application: A Foundation for AI Validation | Pharmaceutical Engineering," ISPE. Accessed: Jun. 24, 2025. [Online]. Available: https://ispe.org/pharmaceutical-engineering/march-april-2022/ai-maturity-model-gxp-application-foundation-ai

[3] European Medicines Agency (EMA), Ed., "Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle_240903." Sep. 09, 2024.
[4] U.S. Food and Drug Administration (U.S. FDA), "Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products." Jan. 2025. Accessed: Jun. 24, 2025. [Online]. Available: https://www.fda.gov/media/184830/download
[5] L. Oestreich, K. Auchincloss, and E. Hartmann, "Trends In FDA FY 2024 Inspection-Based Warning Letters," Pharmaceutical Online. Accessed: May 17, 2025. [Online]. Available: https://www.pharmaceuticalonline.com/doc/trends-in-fda-fy-2024-inspection-based-warning-letters-0001
[6] D. Husman, "Deviation Management - Why Have We Not Solved the Problem?," David Husman Consult. Accessed: May 17, 2025. [Online]. Available: https://www.davidhusmanconsulting.com/post/10-ways-to-engage-with-your-clients
[7] R. D. Snee, "A Systems Approach to Root Cause Analysis and CAPA Investigations," in *ResearchGate*, Philadelphia, PA, 6.12 2017. Accessed: May 17, 2025. [Online]. Available: https://www.researchgate.net/publication/321975850_A_Systems_Approach_to_Root_Cause_Analysis_and_CAPA_Investigations
[8] G. Yenduri *et al.*, "GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions," *IEEE Access*, vol. 12, pp. 54608–54649, 2024, doi: 10.1109/ACCESS.2024.3389497.
[9] N. S. Arden, A. C. Fisher, K. Tyner, L. X. Yu, S. L. Lee, and M. Kopcha, "Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future," *International Journal of Pharmaceutics*, vol. 602, p. 120554, Jun. 2021, doi: 10.1016/j.ijpharm.2021.120554.
[10] J. Kell, "How pharmaceutical companies are training their workers on AI," Business Insider. Accessed: May 18, 2025. [Online]. Available: https://www.businessinsider.com/pharmaceutical-companies-embrace-ai-in-drug-discovery-efforts-2025-3
[11] International Society for Pharmaceutical Engineering (ISPE) GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems (Second Edition), GAMP. Accessed: Sep. 09, 2025. [Online]. Available: https://guidance-docs.ispe.org/doi/book/10.1002/9781946964571
[12] European Medicines Agency (EMA), "Guidance on good manufacturing practice and good distribution practice: Questions and answers | European Medicines Agency (EMA)." Accessed: Mar. 23, 2025. [Online]. Available: https://www.ema.europa.eu/en/human-regulatory-overview/research-development/compliance-research-development/good-manufacturing-practice/guidance-good-manufacturing-practice-good-distribution-practice-questions-answers
[13] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), "Pharmaceutical Quality System Q10." Apr. 06, 2008. doi: 10.1163/ej.9789004163300.i-1081.897.
[14] U.S. Food and Drug Administration (U.S. FDA), "Facts About the Current Good Manufacturing Practice (CGMP)," *FDA*, Aug. 2024, Accessed: Dec. 29, 2024. [Online]. Available: https://www.fda.gov/drugs/pharmaceutical-quality-resources/facts-about-current-good-manufacturing-practice-cgmp