Quartica™

Beyond the Sandbox

Why Real-World Immersion is Essential for AI Authoring Success

Quartica.com

Executive summary

As life sciences organizations explore AI-powered agentic and authoring solutions, many default to sandbox-type evaluations built on synthetic or generic data. While this approach may be suitable for simple productivity tools, it is fundamentally misaligned with the realities of true AI solutions designed to generate complex, health authority-ready documents that must match or exceed the quality of seasoned authors.

This whitepaper explains why traditional sandboxes fail, why synthetic data evaluations are misleading, and how an immersion-based evaluation model provides a more reliable, lower-risk path to production adoption.

The Real Objective of AI Authoring

When evaluating an AI authoring solution, the core question is not whether the software functions in principle, but whether it can:

- Generate documents that approach, match, or exceed human-authored quality
- Adhere to your organization's templates, SOPs, and regulatory expectations
- Reflect your established writing styles and scientific conventions
- Operate reliably within validated, controlled environments

Answering these questions requires the system to be trained on your real data. Synthetic datasets, generic examples, or vendor-curated sandboxes cannot replicate the nuance, variability, and institutional knowledge embedded in actual R&D, regulatory, and safety documents. As a result, conclusions drawn from such evaluations are often meaningless when projected into real-world use.

The Practical Failure of the Sandbox

Beyond data limitations, sandboxes introduce significant operational friction. Sandboxes typically rely on internal users to evaluate unfamiliar software using unfamiliar data, often with limited training. These same users already carry full-time responsibilities, and AI evaluation becomes a secondary task competing for attention.

The result is predictable:

- Superficial testing
- Inconsistent usage
- Feedback driven more by frustration than informed assessment

This is not a failure of the users; it is a structural flaw in the sandbox approach.

A Troubling Industry Signal

Industry data reinforces this reality. According to Lenovo (2025), 88% of paid AI pilots never progress to production.

The reasons are systemic:

- Many AI vendors pursue pilots primarily to secure logos for investor materials
- Pilot fees are often used to fund ongoing product development

- The market is saturated with tools labeled "AI" that offer limited or no true intelligence

Importantly, these failures are rarely due to a lack of AI capability in principle, but rather to evaluation models that fail to reflect real-world operating conditions. For buyers, this creates risk, wasted effort, and evaluation fatigue.

Rethinking AI Evaluation: The Immersion Model

A more effective approach recognizes a simple truth: AI authoring systems must be evaluated in conditions that resemble real use.

A more effective evaluation model recognizes that AI authoring systems must be assessed under conditions that closely resembles real-world use. For authoring of regulated documents, this requires moving beyond the sandbox approach, and toward evaluations that reflect actual data, documents, and operating constraints.

Principles of Immersion-Based Evaluation

- **Your data, not synthetic data:** Models are trained using your documents, templates, SOPs, and writing styles
- **Hands-on, guided evaluation:** A focused two-day workshop pairs your users with our experienced subject matter experts
- **Uncontrolled, realistic usage:** Users work in real authoring scenarios rather than scripted demos
- **Outcome-driven assessment:** You evaluate the quality of actual deliverables, not abstract capabilities

This structure eliminates guesswork and reveals whether the solution can truly perform in your environment.

Lower Risk, Higher Confidence

Immersion-based evaluation provides earlier, more reliable insight into production readiness, allowing organizations to assess document quality, scalability, and governance fit before committing to broader rollout. Instead of funding experimentation or vendor development, organizations gain:

- Clear visibility into document quality
- Direct user feedback grounded in real work

- Confidence in scalability and production readiness
- A faster time-to-value at a lower cost than a full production-based-pilot

Most importantly, decision-makers can determine whether an AI authoring solution is genuinely capable of transforming their processes, or whether it should be eliminated from consideration.

Regulatory Perspective on AI Evaluation

From a regulatory standpoint, the evaluation of AI-assisted authoring systems cannot be separated from the processes and controls under which they operate. Emerging guidance, including the principles articulated by CIOMS Working Group XIV and the risk-based approach of the EU AI Act, reinforces this perspective. These frameworks emphasize human-in-the-loop control, accountability, transparency, and fitness for purpose within real operating contexts. Evaluations conducted on synthetic data or in sandbox environments may demonstrate technical capability but provide limited assurance of regulatory readiness.

As a result, organizations increasingly recognize that meaningful AI evaluation must reflect production conditions, where governance, training, and operational controls are applied consistently and can be inspected and defended.

Conclusion

AI authoring is not a feature to be tested in isolation. In regulated environments, it must be proven under real conditions, with real data, real users, and real governance controls. Sandbox pilots and synthetic datasets obscure the truth, contributing to the industry's high failure rate.

An immersion-based evaluation model replaces speculation with evidence, enabling organizations to move forward with clarity, confidence, and significantly reduced risk.

For organizations serious about AI-driven authoring, immersion is not an alternative to the sandbox, it is the evolution beyond them.

About Quartica

Quartica is a trusted provider of AI-driven solutions purpose-built for regulated life sciences organizations. The Quartica MARS Platform applies domain-trained, enterprise-grade AI to complex clinical, pharmacovigilance, regulatory, and other R&D workflows enabling organizations to generate higher-quality documents faster, with full transparency, validation, and control. Designed to integrate seamlessly with existing systems, data, and SOPs, the Quartica MARS Platform manages workflows to ensure timely delivery and co-authors up to 90% of the first draft of any document across the pharmaceutical enterprise. To learn more about how Quartica and the MARS Platform can support your organization, we invite you to contact us or explore our thought leadership resources.

For more information, please visit www.quartica.com or email intelligence@quartica.com to contact one of our AI for Life Sciences experts.