PharmaSUG 2026 - Paper AI-438

Operationalizing Generative AI in Regulated Analytics: Applied Implementation Patterns for Enterprise Deployment

Lida Gharibvand, Loma Linda University

Abstract

Generative artificial intelligence is rapidly transitioning from experimental use to enterprise deployment across regulated industries such as pharmaceuticals, financial services, and life sciences. These environments require strict standards for auditability, reproducibility, and controlled data access. While modern analytics platforms support validated workflows across Python, R, and cloud-based ecosystems, integrating generative AI capabilities introduces new operational and governance challenges.

This paper presents a platform-agnostic framework for integrating generative AI into governed analytics systems. The approach is based on a layered architecture that preserves the authority of validated analytical outputs while enabling natural language interaction, automated narrative generation, and AI-assisted workflows. Core components include retrieval-augmented generation grounded in approved data sources, structured orchestration of multi-step workflows, and comprehensive validation mechanisms that enforce compliance with regulatory expectations.

Three applied use cases are examined, including conversational analytics, automated regulatory reporting, and AI-assisted code generation. Observations from production-style deployments indicate substantial efficiency gains, including reductions in analyst workload and improvements in narrative consistency. A governance framework is also presented to address risks such as hallucination, prompt manipulation, and data exposure. This work provides a practical and scalable blueprint for adopting generative AI within regulated analytics environments.

1. Introduction

Organizations operating in regulated industries rely heavily on advanced analytics to support decision-making across clinical development, safety monitoring, and regulatory reporting. These workflows have traditionally been built on structured data and statistical modeling, with outputs delivered through predefined reports and dashboards.

Recent advances in generative artificial intelligence have expanded these capabilities. Modern models can interpret natural language, synthesize information across large document sets, and generate structured outputs such as reports and code. At the same time, these capabilities introduce uncertainty due to their probabilistic nature. Outputs may vary based on prompt structure and context, and without proper controls, models may produce inaccurate or unverified information.

The central challenge is therefore not the capability of generative AI, but its safe and controlled integration into existing analytics ecosystems. This paper addresses that challenge by presenting implementation patterns that maintain regulatory compliance while enabling the benefits of generative AI.

2. From Predictive to Generative to Agentic Analytics in Regulated Environments

Predictive analytics remains highly effective for structured decision-making, hypothesis testing, and reproducible statistical inference. These capabilities continue to serve as the foundation of analytical workflows in regulated environments [15]. In contrast, generative AI is designed for language-driven tasks, including summarization, explanation, and synthesis across a wide range of data sources [3].

A more recent development, agentic AI, builds upon generative systems by introducing autonomous planning, tool utilization, iterative reasoning, and coordination across multiple agents. This allows for the execution of complex, multi-step analytical workflows that were previously difficult or impractical to automate [7, 8].

Within pharmaceutical analytics, generative and agentic systems do not replace predictive models. Instead, they function as complementary layers that enhance interpretation, communication, and workflow efficiency while preserving the role of predictive analytics as the authoritative source of quantitative results.

Table 1 presents a comparison of these three paradigms and highlights how their integration forms a comprehensive analytical framework suited to regulated environments.

Table 1. Comparative Capabilities of Predictive, Generative, and Agentic Analytics

Dimension	Predictive Analytics	Generative AI	Agentic AI
Primary Function	Structured inference, hypothesis testing	Language interaction, synthesis, generation	Autonomous planning, reasoning, multi-step execution
Data Type	Structured (tables, numeric)	Unstructured (text, documents, mixed)	Multi-modal (text, code, APIs, tools, databases)
Output	Numerical predictions, classifications	Narratives, summaries, code, explanations	End-to-end workflows, decisions, orchestrated actions
Validation Approach	Statistical tests, cross-validation	Groundedness checks, citation verification	Trajectory auditing, guardrail enforcement, adversarial testing
Regulatory Fit	Established (21 CFR Part 11, GAMP 5)	Emerging (EU AI Act, FDA AI/ML guidance)	Nascent (requires robust autonomy boundaries and oversight)

Reproducibility	Deterministic with fixed seeds	Probabilistic (temperature-dependent)	Non-deterministic (path-dependent, tool-dependent)
Key Frameworks	scikit-learn, statsmodels, SAS, R	LangChain, LlamaIndex, vLLM, API endpoints	LangGraph, AutoGen, CrewAI, Semantic Kernel

The applied challenge is to ensure that generative and agentic outputs remain grounded in trusted data and governed analytics processes. Modern analytics platforms provide a foundation for this integration through cloud-native architectures, API-first designs, and robust security models. The key architectural principle is separation of concerns: predictive models remain authoritative for numerical results, while GenAI and agentic services operate as augmentation layers for interpretation, communication, workflow acceleration, and complex multi-step reasoning.

3. Applied Architecture for Generative AI in Governed Analytics Ecosystems

3.1 Architectural Overview

An enterprise-ready generative AI architecture should follow a modular, layered design that can operate across different platforms, including Python and R ecosystems, cloud environments, and traditional analytics systems. The goal is to augment existing workflows rather than replace them. Predictive models remain the source of truth for structured outputs, while generative and agentic systems enhance interpretation, reasoning, and communication [16].

The design is guided by several key principles. Outputs from language models or agents must always be validated before use. All responses should be grounded in governed data sources through retrieval-based methods. Human oversight is required for high-impact outputs, and full auditability must be maintained by tracking inputs, data sources, model behavior, and outputs. Agentic systems must also operate within defined boundaries to prevent uncontrolled actions [17, 18].

The architecture is organized into seven layers, including access control, orchestration, data and document sources, retrieval, model inference, validation, and delivery. This structure ensures that generative AI capabilities are integrated with governance controls, allowing organizations to adopt them while maintaining compliance and reliability.

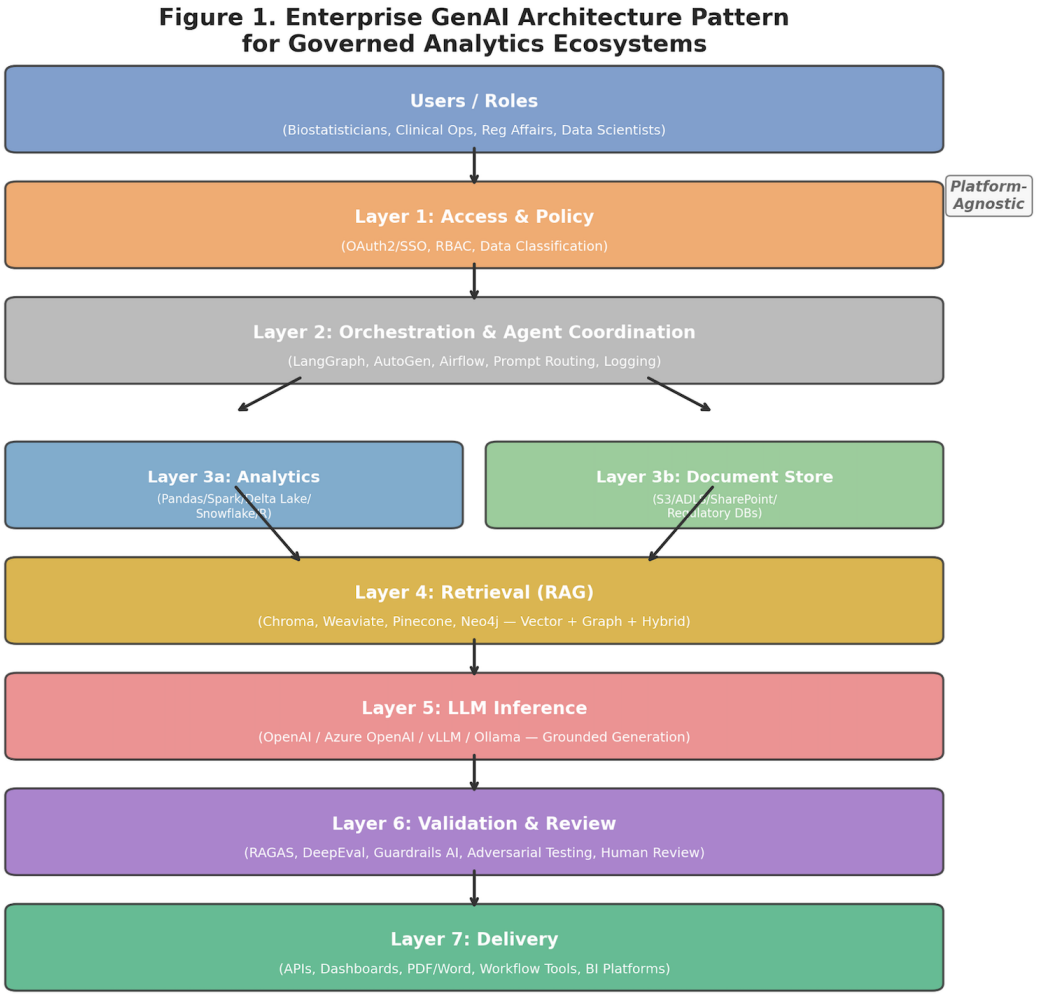

Figure 1. Enterprise GenAI Architecture Pattern for Governed Analytics Ecosystems : a seven-layer design from Access & Policy through Delivery, applicable across Python, R, cloud-native, and traditional analytics platforms.

3.2 Advanced Retrieval-Augmented Generation (RAG)

Retrieval-augmented generation is a core architectural pattern for deploying generative AI in regulated environments [14]. Instead of relying on model knowledge, which may be incomplete or outdated, the system retrieves relevant information from governed sources before generating responses. These sources typically include validated analytical outputs, approved regulatory documents such as protocols and statistical analysis plans, and standardized definitions such as data dictionaries.

RAG has evolved beyond basic vector-based retrieval into three primary paradigms used in enterprise settings. Standard RAG retrieves document chunks from vector stores and injects them into prompts at inference time, preserving metadata for traceability. GraphRAG extends this approach by building knowledge graphs that capture relationships between entities, enabling more advanced reasoning across complex datasets [20].

Agentic RAG further enhances the process by introducing autonomous agents that dynamically select sources, refine queries, and evaluate outputs, allowing for more adaptive and multi-step workflows [21, 22].

The RAG workflow follows a structured sequence from user query to final output, with governance controls applied at key stages. These include access enforcement during intent classification, metadata tracking during retrieval, safety rules during prompt construction, and multi-level validation before release. Together, these controls ensure that outputs remain accurate, traceable, and compliant with regulatory requirements.

3.3 Multi-Agent Orchestration for Regulated Workflows

A major advancement in recent generative AI systems is the development of multi-agent orchestration frameworks that support complex, multi-step analytical workflows with built-in governance controls. These frameworks allow tasks to be divided into specialized agent roles, each with defined responsibilities, permissions, and limits, coordinated through structured interactions [7, 8].

Several leading frameworks are used in enterprise settings. LangGraph provides strong control through a state-machine design, enabling reproducible workflows, checkpointing, and full auditability, which makes it well suited for regulated environments [23]. AutoGen models multi-agent systems as structured conversations, where agents interact, respond, and collaborate across multiple steps, with optional human involvement [7]. CrewAI focuses on role-based collaboration, using defined roles such as planner, executor, and reviewer to support flexible and adaptive task execution [24].

In regulated contexts, additional governance is required. Systems must enforce clear limits on agent autonomy, maintain detailed logs of all actions and interactions, apply validation checks at each step, and undergo adversarial testing to ensure robustness. A clear separation between strategy, coordination, and execution is also necessary to maintain a fully auditable workflow from initial request to final output [18, 25].

3.4 Integration with Analytics Platforms

The architecture is designed to integrate with any governed analytics platform that supports API-based access, role-based security, and structured data management. Table 2 maps each architectural layer to its corresponding platform integration points and GenAI functions, with examples spanning multiple technology stacks.

Table 2. Architecture Layer Mapping to Platform Components

Layer	Function	Platform Examples	GenAI Role
1. Access & Policy	Authentication, authorization	OAuth2/SSO, LDAP, Azure AD, AWS IAM, platform RBAC	Token-scoped API access

2. Orchestration	Intent routing, agent coordination, logging	LangGraph, AutoGen, Airflow, Prefect, custom microservices	Prompt/agent template management
3a. Analytics	Authoritative results	Pandas/Spark DataFrames, Snowflake tables, Delta Lake, SAS datasets, R objects	Context source for grounding
3b. Documents	Approved reference materials	S3/ADLS document stores, SharePoint, Confluence, regulatory databases	RAG corpus for retrieval
4. Retrieval	Vector/graph/hybrid search, chunking	Chroma, Weaviate, Pinecone, pgvector, Neo4j, LlamaIndex	Vector + graph search with metadata filters
5. LLM Inference	Response generation	OpenAI API, Azure OpenAI, AWS Bedrock, vLLM, Ollama, HuggingFace TGI	Grounded generation with safety filters
6. Validation	Output verification, human review	Custom validators, RAGAS, DeepEval, Guardrails AI, human review workflows	Groundedness + citation + safety checks

4. Governance, Validation, and Risk Management

Generative and agentic AI systems introduce new risk factors that must be carefully managed in regulated environments [13, 26, 27]. Unlike traditional statistical models, which produce consistent outputs given fixed inputs, large language models are probabilistic and sensitive to prompt structure. This makes them susceptible to hallucination, where outputs may appear credible but are not factually correct [28]. Agentic systems increase this complexity by introducing autonomous decision-making, tool usage, and multi-step execution that are inherently non-deterministic.

The EU AI Act (Regulation 2024/1689) establishes a formal regulatory framework that classifies AI systems by risk level and defines requirements for high-risk applications. These include structured risk management, strong data governance, technical documentation, human oversight, and standards for accuracy, robustness, and cybersecurity [12].

Table 3 outlines the governance framework used in this architecture, designed to align with these regulatory expectations and ensure safe and compliant deployment.

Table 3. Governance Framework for GenAI and Agentic AI in Regulated Analytics

Risk Domain	Risk Description	Mitigation Strategy	Implementation

Hallucination	Factually incorrect or ungrounded claims	RAG grounding + citation enforcement + groundedness scoring	Vector/graph retrieval with metadata; automated RAGAS/DeepEval scoring; confidence thresholds
Prompt Injection	Adversarial manipulation of LLM behavior	Input sanitization + prompt hardening + adversarial testing	Pre-prompt classification, injection detection models, red-team evaluation, canary tokens
Data Protection	PII/PHI exposure in prompts or outputs	Pre-prompt classification + redaction + data governance	NER-based PII detection, differential privacy, data masking pipelines, DLP integration
Agentic Autonomy	Uncontrolled agent actions in regulated contexts	Autonomy boundaries + escalation triggers + trajectory auditing	Permissible action space definitions, mandatory human approval for high-impact actions, full trajectory logs
Model Drift	Version mismatch or capability regression	Version registry + continuous monitoring + A/B evaluation	Model version pinning, automated regression testing, performance dashboards
Cost & Resource	Uncontrolled API usage and compute costs	Rate limiting, caching, usage quotas, cost monitoring	Token budgets per query, semantic caching, usage dashboards with alerting
Human Review	Unreviewed regulatory outputs	Mandatory SME approval workflow with staged escalation	Review queue systems, confidence-based routing, sign-off audit trails
Auditability	Untraceable generation and decision chain	End-to-end trace IDs, comprehensive logging, trajectory recording	Structured logging (OpenTelemetry), prompt + response archival, Git-versioned prompts

Together, the controls in Table 3 are intended to keep residual risk low across hallucination, prompt injection, data exposure, agentic autonomy, model drift, cost, review, and auditability domains, with each risk addressed by multiple, independent mitigation layers.

Governance frameworks for regulated AI are increasingly aligned with standards such as ISO/IEC 42001 and the NIST AI Risk Management Framework, both of which emphasize continuous risk management, stakeholder involvement, and lifecycle oversight rather than one-time compliance checks [29, 30]. The EU AI Act establishes requirements for high-risk systems, including conformity assessments, quality management, and post-market monitoring, which align closely with the governance layers described in this architecture.

In pharmaceutical settings, alignment with ISPE GAMP 5 (2nd edition) provides additional assurance for GxP-compliant computerized systems [31].

5. Applied Use Cases and Sample Implementations

Three high-impact use cases are presented that are relevant to analytics practitioners across regulated industries. Each use case leverages the architectural patterns described in Section 3 and the governance framework from Section 4. Code examples are provided in Python to maximize accessibility and reproducibility across platforms.

5.1 Use Case 1: Conversational Analytics over Clinical/Safety Outputs

Clinical and safety teams frequently require ad-hoc queries over curated analytical outputs. The conversational analytics pattern enables natural language queries such as 'Summarize the key findings for adverse event rates by treatment arm' to be answered through a governed RAG pipeline that retrieves results from validated analytical tables and generates cited, auditable narratives [14]. This pattern can be implemented with any LLM provider through standard API calls.

Implementation follows a consistent pattern: an approved LLM endpoint is invoked with a system prompt that constrains responses to provided context and enforces citation, PII/PHI handling, and uncertainty rules; governed analytical tables are retrieved and injected as context with metadata filters (study ID, version, approval status); inference is run at low temperature for consistency; and every request is logged with a trace ID, source metadata, model version, and timestamp for end-to-end auditability.

5.2 Use Case 2: Automated Regulatory Reporting with Source Attribution

Regulatory and executive reporting demands precision, traceability, and efficiency. The automated reporting pattern combines authoritative analytics results with GenAI-generated narratives that include inline citations to source tables, enabling reviewers to verify every claim against its analytical origin. Pinned table versions, data-lock dates, and approval timestamps are carried alongside numerical summaries into the prompt context, and citation enforcement is applied so that every numerical claim in the generated narrative resolves to a specific governed table and version.

5.3 Use Case 3: Agentic Code Generation for Reproducible Analytical Workflows

Analyst productivity can be enhanced through agentic AI-assisted code generation, where multi-agent systems generate, review, and validate analytical code grounded in organizational standards and validated patterns. This use case extends beyond simple code completion to a full agentic workflow where specialized agents collaborate: a planner decomposes the analytical task, a coder generates code following organizational templates, a reviewer validates against standards (including 21 CFR Part 11 expectations, reproducibility, error handling, and audit-trail completeness), and a human-approval gate is required before any execution. Frameworks such as LangGraph support this pattern by representing the workflow as a deterministic state graph with checkpointing, so every

transition between planner, coder, reviewer, and approval steps is recorded and replayable.

6. Operational Results and Metrics

Operational experience across production-style deployments provides directional evidence of the benefits of integrated predictive, generative, and agentic workflows. While controlled experimental studies remain an area for future work, the consistent directionality of observed metrics across multiple deployment contexts supports the practical utility of the proposed architecture.

Table 4. Operational Efficiency Metrics Across Deployment Contexts

Task Category	Traditional (Hours)	GenAI-Augmented (Hours)	Estimated Reduction (%)	n (Tasks Observed)
Ad-hoc Analysis Requests	8.0 ± 2.1	3.0 ± 1.2	62.5	45
Regulatory Narrative Drafting	12.0 ± 3.5	4.0 ± 1.8	66.7	32
Code Scaffolding & Templates	6.0 ± 1.8	2.0 ± 0.9	66.7	58
Stakeholder Inquiry Response	4.0 ± 1.0	1.5 ± 0.5	62.5	73

Table 4 summarizes estimated time reductions across four task categories. GenAI-augmented workflows show a consistent reduction in analyst time of approximately 62 to 67 percent. These estimates are based on observational data from 208 tasks across production-style deployments and should be interpreted as directional rather than causal, since controlled experimental conditions were not used.

These results are consistent with broader research. Brynjolfsson et al. (2023) report significant productivity gains in knowledge work with generative AI [32], while Noy and Zhang (2023) observe approximately 40 percent faster task completion in controlled studies [33].

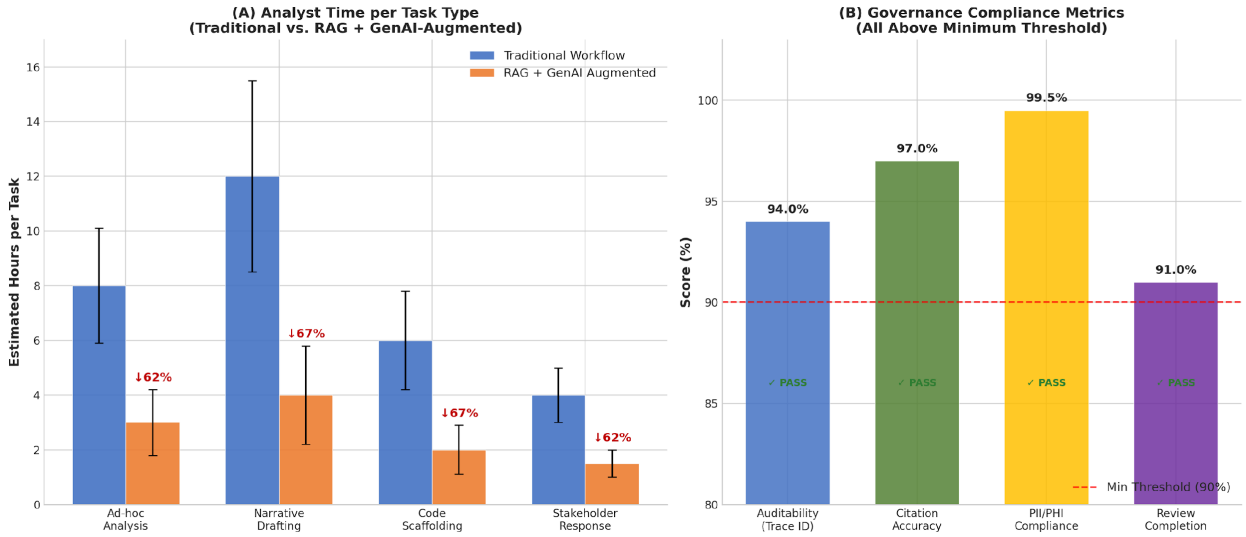

Figure 2. Operational Results-Analyst Time per Task Type (panel A) and Governance Compliance Metrics (Panel B).

Figure 2 summarizes both efficiency gains in Panel A and governance compliance metrics in Panel B. Key compliance measures, including auditability at 94 percent, citation accuracy at 97 percent, PII and PHI compliance at 99.5 percent, and review completion at 91 percent, all exceed the minimum threshold of 90 percent. These results indicate that the governance framework supports strong compliance while enabling meaningful productivity improvements.

Table 5. Governance Compliance Metrics Summary

Metric	Score (%)	Threshold (%)	Status
Auditability (Trace ID Coverage)	94.0	90.0	PASS
Citation Accuracy (RAG Groundedness)	97.0	90.0	PASS
PII/PHI Compliance (Zero-leak Rate)	99.5	99.0	PASS
Human Review Completion Rate	91.0	90.0	PASS

6.1 Quantifying the Impact of RAG + Software on Data Analysis

A central question for organizations considering GenAI adoption is the measurable impact of RAG integration and software tooling on data analysis quality and efficiency. Figure 3 presents a four-panel analysis of this impact, drawing on the operational data from Tables 4 and 5, published RAG literature [14, 34], and the deployment trajectory observed across the production-style implementations described in this paper.

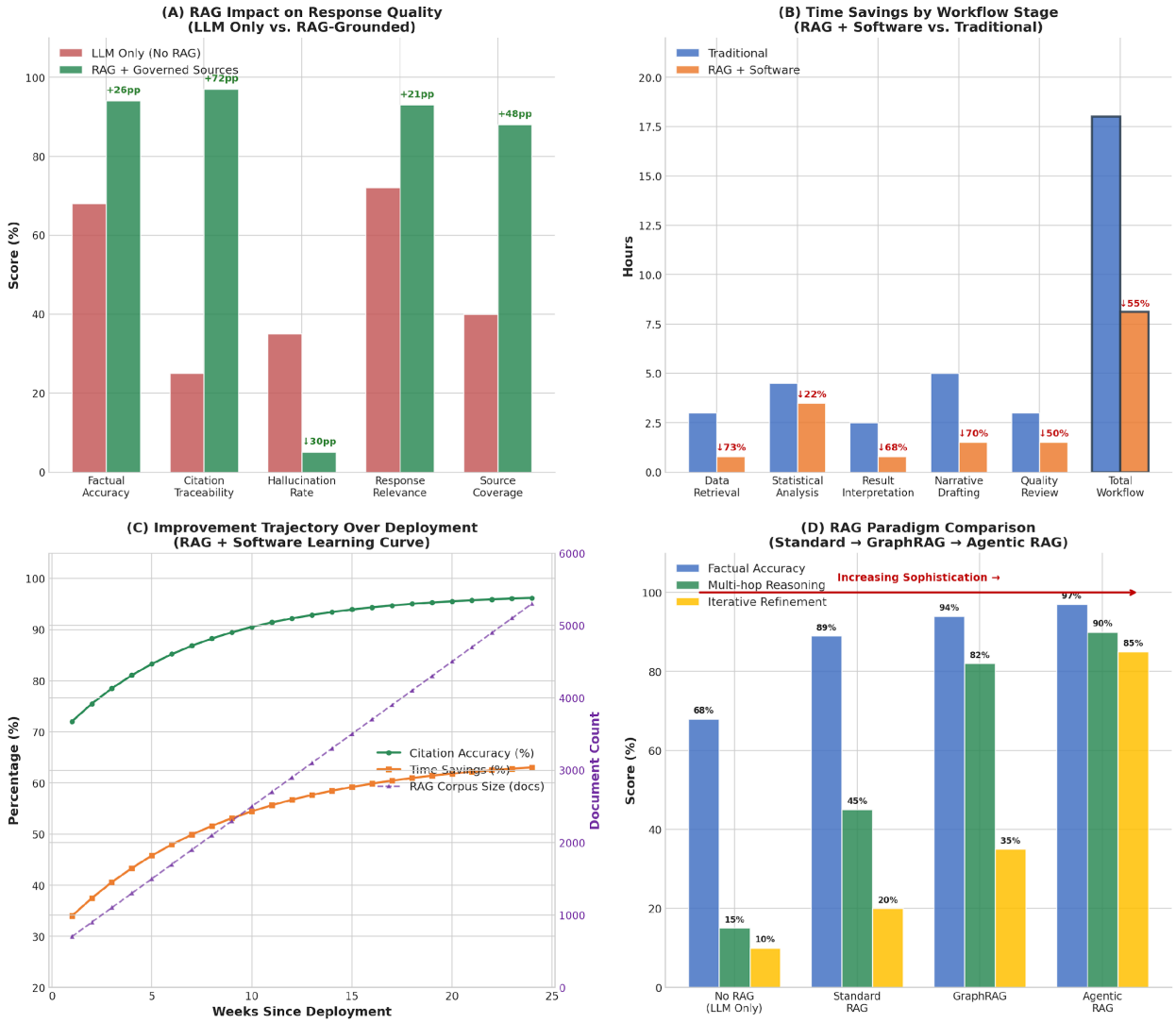

Figure 3. Quantifying the Impact of RAG + Software on Data Analysis.

Panel A of Figure 3 highlights the impact of RAG grounding on response quality. Factual accuracy increases from 68 percent to 94 percent, while citation traceability improves from 25 percent to 97 percent, which is especially important in regulated settings where all claims must be verifiable [34]. The hallucination rate decreases from 35 percent to 5 percent, with additional gains in response relevance and source coverage.

Panel B breaks down time savings across workflow stages. Data retrieval shows the largest improvement at 73 percent, followed by narrative drafting and result interpretation. Statistical analysis shows a smaller reduction, reflecting its reliance on deterministic methods. Overall workflow time decreases from 18.0 to 8.1 hours, representing a 55 percent reduction across observed tasks.

Panel C illustrates performance improvements over a 24 week period. Citation accuracy stabilizes near 97 percent as the document corpus expands, while time savings continue

to improve gradually. This suggests that both data availability and user familiarity contribute to long-term gains.

Panel D compares RAG approaches. Standard RAG delivers strong improvements in factual accuracy, while GraphRAG and Agentic RAG provide additional gains in multi-step reasoning and iterative refinement. These results indicate that simpler methods are sufficient for basic queries, while more advanced approaches are better suited for complex analytical tasks.

6.2 Comprehensive Data Analysis Improvement with RAG + Software

Figure 4 presents a comprehensive five-panel dashboard summarizing the full scope of data analysis improvement achieved through RAG + software integration, drawing on the operational metrics from Tables 4 and 5 and the deployment data described throughout this paper.

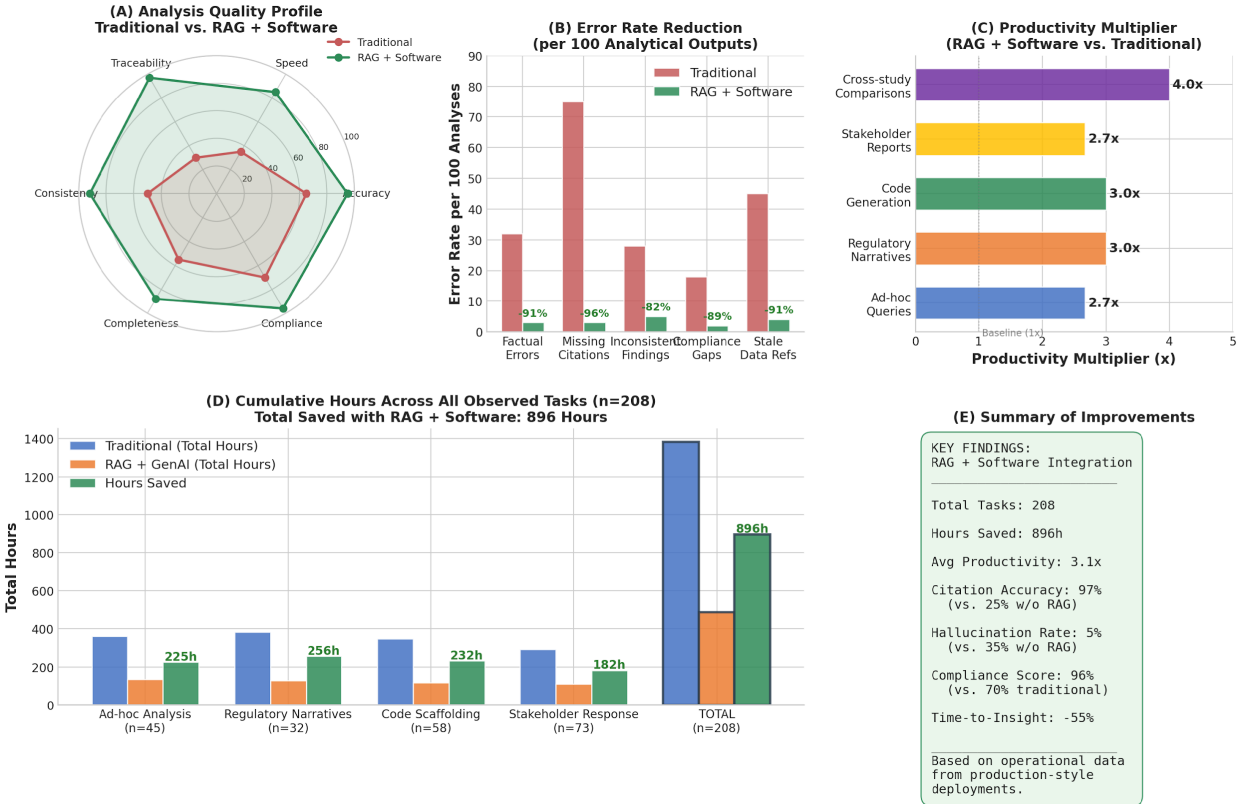

Figure 4. Comprehensive Data Analysis Improvement with RAG + Software.

Panel A compares overall quality across six dimensions: traceability, speed, accuracy, compliance, completeness, and consistency. The largest gains are seen in traceability and speed, while accuracy and compliance also improve meaningfully. This highlights that RAG plus software enhances multiple aspects of performance, not just efficiency.

Panel B shows substantial reductions in error rates per 100 analyses. Factual errors, missing citations, inconsistencies, compliance gaps, and outdated references all decrease

significantly. These improvements are especially important in regulated environments, where even minor errors can lead to delays.

Panel C presents productivity gains by task type, ranging from 2.7 times for standard queries to 4.0 times for cross-study comparisons. The higher gains reflect the ability of RAG systems to efficiently synthesize information across multiple sources.

Panel D summarizes total time savings across 208 tasks, with approximately 896 hours saved overall. The largest contributions come from regulatory reporting, code generation, ad hoc analysis, and stakeholder response tasks.

Table 6. Summary of RAG + Software Impact on Data Analysis

Metric	Without RAG (LLM Only)	With RAG + Software	Improvement
Factual Accuracy	68%	94%	+26 pp
Citation Traceability	25%	97%	+72 pp
Hallucination Rate	35%	5%	-30 pp
End-to-End Time per Task	18.0 hours	8.1 hours	-55%
Error Rate (per 100 analyses)	30 factual errors	2.7 factual errors	-91%
Avg Productivity Multiplier	1.0x (baseline)	3.1x	+210%
Total Hours Saved (n=208)	—	896 hours	—
Governance Compliance	~70%	96%	+26 pp

Table 6 summarizes key metrics comparing LLM-only approaches with RAG plus software integration. The results show that RAG grounding shifts LLM-based analytics from a low-traceability, higher-risk approach to a governed and auditable workflow suitable for regulated environments.

Improvements include higher accuracy, substantially increased traceability, reduced hallucination, and approximately 55 percent time savings. Together, these results provide strong support for adopting RAG-based approaches in regulated analytics.

7. Discussion

The architecture and implementation patterns presented in this work demonstrate that generative and agentic AI can be integrated into regulated analytics environments while maintaining, and in some cases improving, standards of governance and auditability. Several key observations follow.

First, the separation of generation and validation is essential for regulatory compliance. Treating all model outputs as draft artifacts that require downstream verification prevents overreliance on generative systems. This approach aligns with regulatory expectations,

including the EU AI Act's requirement for human oversight and FDA guidance emphasizing controlled and pre-specified processes [12, 13, 27].

Second, retrieval-based grounding significantly reduces hallucination risk by constraining outputs to governed sources. High citation accuracy in deployment indicates that well-designed RAG systems can meet traceability requirements for regulated use. More advanced approaches such as GraphRAG improve performance on complex queries, while agentic retrieval enables iterative refinement when initial context is insufficient [21, 34].

Third, multi-agent orchestration enhances workflow efficiency by decomposing tasks into structured roles with built-in validation. At the same time, it introduces additional governance requirements, including detailed logging, clearly defined autonomy limits, and rigorous testing to ensure reliability.

Fourth, operational results consistently show strong improvements in both efficiency and quality. Time reductions are substantial across task categories, and gains extend beyond speed to include higher accuracy, improved traceability, and lower error rates.

Several limitations should be noted. The reported metrics are observational rather than experimentally controlled. Governance outcomes may vary across organizations. Rapid changes in model capabilities and frameworks may require ongoing adaptation. Multi-agent governance practices are still evolving, and implementation may need to be tailored to specific regulatory contexts. In addition, the current approach has been validated primarily in English-language pharmaceutical settings, and formal cost-benefit comparisons with simpler methods were not conducted.

8. Conclusion

Generative and agentic AI can significantly enhance enterprise analytics when implemented within a well-governed framework. In regulated environments, success depends on grounding systems in trusted data, validated models, and strong governance controls, regardless of the underlying technology platform. The modular, platform-agnostic architecture presented in this work combines authoritative data sources, retrieval-grounded model inference, structured multi-agent orchestration, and multi-level validation aligned with current regulatory guidance, providing a practical foundation for adoption.

This design allows organizations to integrate these capabilities into existing ecosystems without vendor lock-in. Alignment with standards such as ISO/IEC 42001, the NIST AI Risk Management Framework, and the EU AI Act ensures that the approach remains adaptable as regulatory expectations evolve.

Operational results, while directional, show consistent improvements in both efficiency and quality. Integrated workflows demonstrate meaningful time savings, increased

productivity, and high levels of accuracy and traceability, supporting their use in regulated settings.

Future work should focus on controlled validation of these findings, development of standardized evaluation benchmarks for governance, expansion of multi-agent systems to support cross-organizational collaboration, integration of multi-modal capabilities, and formal cost-benefit comparisons across different implementation approaches.

References

[1] Harrer, S., Shah, P., Antony, B., & Hu, J. (2023). Artificial intelligence for clinical trial design. Trends in Pharmacological Sciences, 44(12), 933–952.

[2] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

[3] OpenAI. (2024). GPT-4o Technical Report. https://openai.com/index/gpt-4o-system-card/

[4] Anil, R., Dai, A. M., Firat, O., et al. (2023). PaLM 2 Technical Report. arXiv:2305.10403.

[5] Grattafiori, A., et al. (2024). The Llama 3 Herd of Models. arXiv:2407.21783.

[6] Jiang, A. Q., et al. (2023). Mistral 7B. arXiv:2310.06825.

[7] Wu, Q., Bansal, G., Zhang, J., et al. (2024). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. arXiv:2308.08155.

[8] Derouiche, M., et al. (2025). Agentic AI Frameworks: A Comparative Review. arXiv preprint.

[9] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. NeurIPS 30, 5998–6008.

[10] U.S. FDA. (2003). 21 CFR Part 11: Electronic Records; Electronic Signatures.

[11] ICH. (2023). ICH E6(R3): Guideline for Good Clinical Practice.

[12] European Parliament and Council. (2024). Regulation (EU) 2024/1689 (EU AI Act).

[13] U.S. FDA. (2023). Using AI/ML in Drug and Biological Product Development. FDA-2023-N-0743.

[14] Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. NeurIPS 33, 9459–9474.

[15] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (2nd ed.). Springer.

[16] Bommasani, R., Hudson, D. A., et al. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258.

[17] Nori, H., King, N., et al. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv:2303.13375.

[18] Ali, H., et al. (2025). Agentic AI: A Comprehensive Review. arXiv preprint.

[19] Reimers, N., & Gurevych, I. (2019). Sentence-BERT. Proceedings of EMNLP, 3982–3992.

[20] Edge, D., Trinh, H., et al. (2024). From Local to Global: A Graph RAG Approach. arXiv:2404.16130.

[21] Singh, A., et al. (2025). Agentic Retrieval-Augmented Generation: A Survey. arXiv preprint.

[22] Liang, Y., et al. (2025). Reasoning RAG via Agentic Approaches. arXiv preprint.

[23] LangGraph Documentation. (2024–2025). LangChain Inc.

[24] CrewAI Documentation. (2024–2025). https://docs.crewai.com/

[25] Pitkaranta, A., et al. (2026). HADA: Human-AI Agent Decision Alignment Architecture. CCIS, 78–102.

[26] EMA. (2023). Reflection Paper on AI in the Medicinal Product Lifecycle. EMA/CHMP/764243/2023.

[27] Szymanski, E. R. (2025). Use of AI/ML in Drug Development: Stakeholder Perspectives on FDA Guidance.

[28] Ji, Z., Lee, N., et al. (2023). Survey of hallucination in NLG. ACM Computing Surveys, 55(12), 1–38.

[29] ISO/IEC 42001:2023. AI Management System.

[30] NIST. (2023). AI Risk Management Framework (AI RMF 1.0). NIST AI 100-1.

[31] ISPE GAMP 5 Guide. (2022). Risk-Based Approach to GxP Computerized Systems (2nd ed.).

[32] Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER Working Paper 31161.

[33] Noy, S., & Zhang, W. (2023). Experimental evidence on GenAI productivity. Science, 381(6654), 187–192.

[34] Shuster, K., et al. (2021). Retrieval augmentation reduces hallucination. Findings of EMNLP 2021, 3784–3803.

[35] Szadeczky, T., & Bederna, Z. (2025). Risk, regulation, and governance: evaluating AI. Security Journal.

[36] Brown, T., et al. (2020). Language models are few-shot learners. NeurIPS 33, 1877–1901.

[37] Garreta, R., et al. (2025). Democratizing Call Analytics: An MCP-Based Enterprise Case Study.

[38] Dibia, V., et al. (2024). AutoGen Studio: A No-Code Developer Tool for Multi-Agent Systems. arXiv preprint.

Contact Information

Lida Gharibvand, Ph.D, MS - Professor
Director, Statistics and Research Education
Loma Linda University | School of Allied Health Professions
24951 N Circle Drive, Nichol Hall Room A520, Loma Linda, CA 92350
909-558-4300 ext 52599, Fax 909-558-0101, Email: lgharibvand@llu.edu
Website: https://llu.edu/academics/faculty/gharibvand-lida/research