Enhancing Compliance, Security & Research in Pharma with Self-Hosted LLM & RAG.

CODE IT

Pharmaceutical companies handle sensitive clinical, regulatory, and research data under strict standards such as GxP, GDPR, HIPAA, and FDA 21 CFR Part 11.

Third-party AI tools pose data privacy risks, making them unsuitable for confidential information.

To correspond to the regulations and processes the pharmaceutical research companies rely on:

— Electronic Document Management Systems (EDMS)

— Quality Management Systems (QMS)

— Regulatory Information Management Systems (RIMS)

While these systems improve document storage and approvals, they still require manual data processing, verification, and extraction, consuming significant employee time. Studies show:

— Analytical reports show EDMS still demands significant effort for document processing, routing, and comparison, using **up to 2.5x more resources than AI-driven solutions.**

— **Up to 40% of employee time is spent** searching for documents, highlighting inefficiencies in traditional systems **without AI**.

— Astera's insights confirm **AI document processing's *"hours to seconds"* benefit**, speeding up retrieval and boosting productivity, as backed by research and case studies.

CODE IT

Key Challenges & Solution.

/01/

Time-consuming manual review

Compliance teams still spend hours reading through regulatory documents and reports to ensure adherence to GxP, FDA 21 CFR Part 11, and HIPAA.

/02/

Inefficient search and retrieval

Traditional systems rely on metadata and keyword searches, often failing to extract meaningful insights from complex unstructured data.

/03/

High risk of human errors

Manual processing leads to delays, inconsistencies, and potential compliance violations.

/04/

Overburdened regulatory and QA teams

Employees spend excessive time on document validation instead of focusing on high-value analytical tasks.

SOLUTION

Self-Hosted LLM & RAG

Deploying an on-premise, self-hosted RAG system with LLM enables pharma organizations to use AI insights while controlling data and meeting strict compliance and security standards.

CODE IT

Challenges in compliance and data security.

Pharmaceutical organizations deal with compliance challenges due to sensitive data.

Traditional document management and search methods create inefficiencies and risks:

High cost of compliance management

Ensuring adherence to industry regulations requires manual effort and time, delaying research and decision-making.

Regulatory documentation complexity

With thousands of scientific reports, clinical trials, and guidelines, retrieving relevant insights is slow and error-prone.

Risk of data exposure

3rd party AI solutions do not offer sufficient protection for proprietary research and patient data, leading to potential regulatory violations.

Cybersecurity threats

Data breaches and unauthorized access pose risks to patient privacy and corporate intellectual property.

CODE IT

Security architecture of the self-hosted RAG system.

Deploying an on-premise AI solution ensures that security and governance policies align with pharmaceutical industry requirements.

The system provides end-to-end data protection with the following measures:

Real-time logging & monitoring Tracks all interactions with the system to meet auditability requirements.	**Secure API gateways** AI model interactions are protected with OAuth 2.0 authentication and rate limiting to prevent unauthorized access.	**Data masking & anonymization** Sensitive information (e.g., patient records) is automatically anonymized before processing.
Role-based access control (RBAC) Only authorized personnel can access, query, or modify sensitive data.	**Regular penetration testing & security audits** Routine assessments identify vulnerabilities and maintain compliance.	**Multi-factor authentication (MFA)** Strengthens access control to prevent unauthorized system usage.

CODE IT

Proposed solution.

Retrieval-augmented generation (RAG) powered by a self-hosted LLM.

01. Integration with internal data repositories

Ingesting and annotating internal documents, publications, clinical reports, and regulatory papers.

02. User query processing

A RAG model processes user queries by retrieving relevant information from a vectorized document index (e.g., FAISS, Weaviate, Milvus).

03. Large language model (LLM) integration

An LLM (Mistral, Llama, and similar) enriches responses, providing context-aware summaries and evidence-based insights.

04. Compliance enabled by security modules

Source validation, response accuracy monitoring, and full query logging for auditability.

Benefits

- Reduction in human regulatory document search time by 60% (McKinsey).
- Enhanced compliance efficiency through automated document validation.
- Elimination of data leakage risks by keeping AI processing within a secured infrastructure.
- Accelerated drug development cycles due to faster research workflows.
- Cost savings thanks to reduced reliance on external compliance consultants and expedites regulatory approvals.

Return on Investment (ROI)

- Lower compliance costs by automating regulatory document handling.
- Reduced risk exposure from data breaches and regulatory violations.
- Faster knowledge retrieval for pharmaceutical R&D and legal teams.

CODE IT

Risk Mitigation Strategy.

Potential AI misinterpretation	→	Implement human-in-the-loop verification for high-risk queries
Infrastructure costs	→	Optimize AI workloads with on-premises GPU acceleration to balance cost and performance.
Regulatory updates	→	A system is designed for continuous learning, ensuring real-time compliance updates from regulatory authorities.

Estimated implementation.

Data preparation & PoC (1–2 months)

Secure document ingestion & indexing.

Model development (2–4 months)

Training & fine-tuning for compliance data.

Security & compliance testing (4–5 months)

Penetration testing & regulatory validation.

Pilot testing (5–6 months)

Real-world validation within a pharmaceutical setting.

Full deployment (6+ months)

Company-wide rollout with compliance oversight.

CODE IT

Conclusion

Self-hosted AI is the only fully compliant and secure solution for pharmaceutical firms. Deploying retrieval-augmented generation (RAG) with a self-hosted LLM ensures strict adherence to industry regulations, enhanced security, and improved compliance efficiency.

By investing in an on-prem AI solution, pharmaceutical companies can safeguard their intellectual property, ensure regulatory approval, and accelerate scientific innovation.

CODE IT

Next Steps: Unlock AI Power in Your Business

Ready to transform your business with AI? Book a free consultation with a CodeIT AI specialist to:

- Discover how AI can transform your business
- Discuss unique solutions to implement
- Get a custom AI implementation strategy

Contact us to discuss your AI strategy!

sales@codeit.us

+1 (716) 342 11 33