78
/ 100

Good Fit

A governed knowledge-retrieval assistant can materially accelerate the medical information desk while preserving regulatory control.

BUSINESS OVERVIEW

Industry Context

INDUSTRY
Pharmaceuticals

CHARACTERISTICS CONSIDERED
- All HCP-facing communication is subject to medical, legal, and regulatory review.
- Pharmacovigilance obligations apply to every enquiry channel.
- Approved content (labels, CDS, PV-cleared response documents) defines the boundary of allowable claims.
- Inspectors expect full traceability of who said what, on which evidence, with which version.

KEY ASSESSMENT IMPLICATIONS
- Architecture is retrieval-grounded with citation enforcement — not free generation.
- Reviewer gating is non-negotiable on every released response.
- PV signal routing is a first-class workflow concern, not an afterthought.

INDUSTRY-SPECIFIC RISKS
- Off-label content reaching an HCP creates a regulatory event.
- Missed PV signal at intake creates patient-safety and compliance risk.
- Stale corpus following a label update propagates incorrect responses.

INDUSTRY GOVERNANCE CONSIDERATIONS
- Medical-affairs sign-off on workflow design and on the approved corpus.
- Quarterly content-and-model audit with documented sampling.
- Change-control before any expansion beyond the Phase 1 product family.

INDUSTRY VALUE DRIVERS
- Response throughput and turnaround
- Reviewer productivity
- Consistency of HCP communication
- Inspection-readiness
- PV signal capture quality

Executive Summary

A governed knowledge-retrieval assistant can materially accelerate the medical information desk while preserving regulatory control.

INVESTMENT SIGNAL
Strong — proceed

TIME TO VALUE
12–16 weeks for a controlled launch on a single product family

FIT SCORE
78 / 100 — Good Fit

WHY NOW
- Approved content is already digital and reasonably well-structured, making RAG feasible.
- Regulatory expectations on traceability align well with RAG citation enforcement.
- Enquiry volume is outpacing the team's ability to scale linearly with headcount.

SITUATION
A mid-size pharmaceutical company's Medical Information team handles ~3,500 healthcare-professional (HCP) enquiries per month across 14 marketed products. Responses must be drawn only from approved labels, core data sheets, and PV-cleared response documents. Median response time is 36 hours; the team is struggling to keep up with launches and regulatory cycles.

RECOMMENDATION
Deploy a governed retrieval-augmented (RAG) assistant that drafts responses to inbound HCP enquiries strictly from an approved content corpus, with mandatory medical-affairs review prior to release, full citation traceability, and PV signal-routing on suspected adverse events.

TOP VALUE DRIVER — RESPONSE THROUGHPUT AND TURNAROUND
Cutting median response time from 36 hours to under 6 hours, with capacity to absorb 30–40% more enquiry volume without additional headcount.

KEY RISK — OFF-LABEL OR FABRICATED CONTENT REACHING HCPS
Any response that references information outside the approved corpus would create a regulatory event. Mitigated by retrieval-only generation, mandatory medical-affairs review, and citation enforcement.

Assessment Confidence

CONFIDENCE — HIGH
Use case is well-bounded, content corpus is identifiable, and reviewer-gated workflows are standard in this industry.

Why This Recommendation Was Generated

- Task benefits from LLM language fluency over a curated, retrievable corpus.
- Human-irreplaceable judgement is preserved through reviewer gating, capping autonomy at 2.
- Score capped at 78 because PV obligations and label-update freshness add residual risk.
- ROI is concrete: throughput, capacity, and reviewer productivity are quantified.
- Risk envelope is bounded by mandatory reviewer release on every response.

Assessment Assumptions

- Approved content is digital, versioned, and accessible via API.
- Medical reviewers have capacity to review drafts at projected enquiry volumes.
- PV intake exposes an API or queue for routed suspected adverse events.
- Provider BAA and EU data residency are achievable with the selected platform.
- Inspectors will accept citation chains stored in the workflow's audit log as evidence.
- Phase 1 scope is limited to a single product family.

Opportunity Sizing

VERDICT — APPROPRIATE SCOPE

Phase 1 is bounded to a single product family with reviewer gating, with measurable ROI in 12–16 weeks.

Information That Would Improve This Assessment

- Sample of 100 representative enquiries to test retrieval and PV detection performance.
- Current PV intake API maturity and latency.
- Frequency and channel of label updates per product.
- Reviewer calibration baseline for accept/edit/reject behaviour.

Assessment Limitations

This assessment is based on the information provided and generated using Bodhvega's structured evaluation framework. Results may vary depending on industry-specific requirements, regulatory constraints, organizational maturity, data quality and availability, existing technology landscape, and business operating model. This assessment should be used as decision-support guidance and not as a substitute for detailed business, architectural, legal, or regulatory review.

AI OPPORTUNITIES

Business Value Map

Concentrated value in response throughput, medical reviewer productivity, and consistency of HCP communication, with meaningful secondary upside in PV signal routing.

Customer Experience Significant

Customer Experience

Faster turnaround strengthens HCP relationships and frees reviewers for complex cases. Industry benchmarks support a 5–6x improvement.

Timeframe: 0–4 months

Labor Productivity Significant

Productivity Improvement

At ~$95K loaded per medical reviewer FTE, a 40% lift on 8 FTE = ~$300K of redeployable capacity.

Timeframe: 0–6 months

Quality & Consistency Significant

Risk Reduction

Reduces variance in HCP communications and audit risk; supports inspection-readiness.

Timeframe: 0–4 months

Compliance Moderate

Compliance Improvement

Reduces risk of missed PV signals at intake and improves time-to-PV-case-open.

Timeframe: 3–6 months

Scalability Moderate

Scalability Enablement

Avoids ~$190K of incremental hiring across the next two launch cycles.

Timeframe: 6–12 months

Audit Readiness Moderate

Compliance Improvement

Citation chain reduces audit preparation effort and supports inspector queries.

Timeframe: 3–9 months

Knowledge Management Moderate

Operational Efficiency

Curated answer library compounds value as corpus grows.

Timeframe: 6–12 months

Customer Satisfaction Moderate

Customer Experience

Realistic 5–10 point lift in HCP satisfaction; reinforces medical brand equity.

Timeframe: 6–12 months

PRIORITY RECOMMENDATION

Priority Recommendation

RECOMMENDATION

Hybrid

Drafting from an approved corpus benefits from LLM language fluency, but release must remain a deterministic, reviewer-gated workflow with PV routing. A hybrid pattern captures the throughput gain without exposing the company to off-label or unapproved content reaching HCPs.

Is the task multi-step with branching logic?

Yes

Intake !' classification !' retrieval !' drafting !' PV check !' reviewer release is mult

Are the inputs structured and predictable?

Partially

HCP enquiries vary in clarity; approved source content is well-structured and curated.

Are mistakes recoverable?

Partially

Pre-release errors are recoverable; a released off-label statement creates a regulatory event.

Is the value of automation high relative to risk?

Yes

Clear throughput and capacity benefits with mandatory reviewer release bounding the residual risk.

GOLDILOCKS CHECK — APPROPRIATELY SCOPED

Phase 1 limited to routine enquiries on a single product family with full reviewer gating — right-sized for a 12–16 week launch.

IMPLEMENTATION OPTIONS

Platform Comparison

Azure OpenAI + Azure AI Search Recommended — Suitability 86/100

Monthly cost: $1,800–$3,500 (GPT-4 class + managed vector search) · Scalability 5/5 · Robustness 5/5

Strong regulated-industry fit: BAA available, EU data residency, mature managed RAG building blocks. Pros: tight Microsoft 365 integration, enterprise governance. Cons: requires Azure footprint. Best for: regulated enterprises already on Microsoft. Time to deploy: 8–12 weeks.

RECOMMENDED BECAUSE

- Existing pharma compliance footprint on Microsoft 365 reduces vendor onboarding effort.
- Managed RAG components shorten time to a controlled Phase 1 launch.
- Regional data residency supports EU and APAC affiliate rollouts.

AWS Bedrock + Kendra

Suitability 80/100

Monthly cost: $1,500–$3,200 (Bedrock + Kendra Enterprise) · Scalability 5/5 · Robustness 5/5

Strong for AWS-native shops; Kendra brings enterprise-grade retrieval. Pros: VPC isolation, BAA, model choice. Cons: requires AWS footprint and integration build-out. Best for: AWS-native pharma operations. Time to deploy: 10–14 weeks.

RECOMMENDED BECAUSE

- Strong choice when existing GxP systems already run on AWS.
- Kendra's tuned retrieval suits dense regulatory corpora.

Google Vertex AI + Vertex AI Search

Suitability 78/100

Monthly cost: $1,400–$3,000 (Gemini + Vertex Search committed-use) · Scalability 5/5 · Robustness 5/5

Strong multimodal RAG with enterprise SLAs. Pros: Gemini retrieval grounding, committed-use pricing, EU residency. Cons: requires GCP footprint. Best for: Google Cloud ecosystem deployments. Time to deploy: 8–12 weeks.

RECOMMENDED BECAUSE

- Vertex AI Search has strong document grounding suitable for regulated content.
- Committed-use pricing supports predictable per-response cost modelling.

Claude (Anthropic) + custom retrieval

Suitability 74/100

Monthly cost: $900–$2,200 (Anthropic API) · Scalability 4/5 · Robustness 5/5

Excellent reasoning and low hallucination rates on structured retrieval. Pros: high-quality drafting. Cons: retrieval layer must be built separately; provider concentration. Best for: teams prioritising response quality with in-house retrieval. Time to deploy: 10–14 weeks.

RECOMMENDED BECAUSE

- Strong instruction-following and low fabrication rates suit citation-backed drafting.

n8n + hosted LLM

Suitability 60/100

Monthly cost: $300–$900 (self-hosted) + LLM API · Scalability 3/5 · Robustness 3/5

Useful for orchestrating the workflow tier in early prototypes. Pros: low cost, visual iteration. Cons: weaker enterprise controls for regulated workloads. Best for: early prototyping only. Time to deploy: 3–6 weeks.

RECOMMENDED BECAUSE

- Acceptable for prototyping; not a production target for regulated workflows.

ACTION PLAN

Blueprint Canvas

TASK / GOAL

Produce a reviewer-ready, fully cited draft response to an inbound HCP medical information enquiry within 30 minutes, restricted to the approved content corpus.

TOOLS

Approved-content vector index (labels, CDS, PV-cleared response library), Classifier for enquiry type and product, PV signal detector, Reviewer workbench with citation viewer, Audit log store

ENVIRONMENT

Internal medical-affairs workbench with SSO, role-based access, and full activity logging.

FEEDBACK LOOP

Reviewer accept / edit / reject decisions and the diff between draft and released response are captured weekly to improve retrieval quality and prompt heuristics.

AUTONOMY LEVEL
2 / 5

WHEN TO STOP
Escalate to a senior reviewer when retrieval confidence is low, when the enquiry is outside the approved corpus, when PV indicators are detected, or when the enquiry concerns a special population (paediatric, pregnancy, hepatic/renal impairment).

RISKS & EDGE CASES
- Hallucination or extrapolation beyond approved content
- Missed PV signal at intake
- Stale corpus following a label update
- Citation drift if source IDs are not version-controlled

Rollout, Oversight & Success Measures

OBJECTIVE
Draft an accurate, fully cited response to an HCP enquiry using only the approved content corpus, ready for medical-affairs reviewer release within 30 minutes.

TOOLS
Approved-content vector index, enquiry classifier, PV signal detector, reviewer workbench, audit log store.

APPROACH
1. Classify enquiry (product, topic). 2. Retrieve approved content. 3. Run PV check. 4. Draft response with inline citations. 5. Emit to reviewer queue with confidence scores.

CONSTRAINTS
Never generate content beyond retrieved approved sources. Always cite. Always route suspected adverse events to PV. Never release without reviewer approval.

SUCCESS
"e85% of routine drafts accepted by reviewers with "d2 edits; zero released r
content.

QUALITY CHECKLIST
- ' Workflow goal is concrete and measurable
- ' Reviewer approval is mandatory before release
- ' PV signal routing is explicit
- ' Citation enforcement is built into the prompt
- ' Corpus version and confidence are recorded in audit log

GOVERNANCE

Guardrails & Governance Controls

Overall Risk: High
Operating in a regulated environment requires strict retrieval grounding, mandatory reviewer release, PV signal routing, and a versioned content corpus. Residual risk is medium under these controls.

HUMAN OVERSIGHT REQUIRED
Yes — All HCP-facing communication must remain under qualified medical reviewer control.

REGULATORY FLAGS
GDPR, EU AI Act, HIPAA

Hallucination Risk High

Off-label or fabricated content in an HCP-facing response would create a regulatory event.

MITIGATION

Retrieval-only generation, citation enforcement, mandatory reviewer release, weekly quality sampling.

Regulatory Compliance High

Pharmacovigilance obligations require timely capture of suspected adverse events at intake.

MITIGATION

Automated PV signal detection at intake with mandatory routing to PV intake and senior reviewer.

Human-in-the-Loop High

No HCP-facing response may be released without a qualified medical reviewer's approval.

MITIGATION

Workflow blocks release until reviewer approves; approver identity logged.

Audit Logging Medium

Inspectors require full traceability of which source documents and corpus version informed each response.

MITIGATION

Per-response capture of source IDs, corpus version, model version, retrieval confidence, and PV flag.

Data Privacy Medium

Enquiry content may include patient-identifiable information shared by HCPs.

MITIGATION

PII redaction at intake; no training on enquiry content; provider BAA in place.

TECHNICAL DETAIL

Full Deployable Prompt

```
Workflow Goal
Draft a citation-backed response to an inbound HCP enquiry from the approved content corpus,
ready for medical-affairs reviewer release within 30 minutes.

Systems & Integrations
- Approved-content vector index (labels, core data sheets, PV-cleared response library)
- Enquiry classification service
- PV signal detection service
- Reviewer workbench (citation viewer, accept/edit/reject)
- Audit log store
- Version-controlled corpus repository

Workflow Logic
1. Ingest enquiry; classify product and topic.
2. Run PV signal detector. If detected, route to PV intake and flag the draft for senior
reviewer.
3. Retrieve top-k approved passages with similarity and source IDs.
4. Draft response strictly grounded in retrieved passages; every claim cites at least one
source.
5. Attach corpus version, retrieval confidence, and PV flag to the draft record.
6. Place in reviewer queue.

Data Requirements
Approved labels, core data sheets, PV-cleared response documents, historical Q&A archive,
classification taxonomy, PV intake API.

Rollout Phases
Phase 1 (weeks 1–8): build retrieval + drafting on one product family with reviewer gating.
Phase 2 (weeks 9–14): expand to portfolio + PV routing in production. Phase 3 (weeks 15–20):
instrument reviewer-edit feedback into retrieval tuning.

Governance & Oversight
```

Mandatory medical-affairs reviewer approval before any release. Quarterly content-and-model audit. Inspection-ready citation chain on every released response.

Operational Readiness
Train reviewers on the workbench and on calibration of accept/edit/reject. Define rollback to fully manual handling if accept-rate < 60% for two consecutive weeks.

Implementation Complexity
Medium. Highest risk areas are corpus freshness following label updates and consistency of citation quality.