

INDUSTRY BRIEF · FOR CIO · CMIO · CHIEF DIGITAL · HEAD OF AI · CISO

CLINICAL AI. PHI- SOVEREIGN. AUDIT-READY.

Clinical AI lives or dies at the PHI boundary. XERJ is a self-hosted data plane that holds logs, embeddings, and agent memory inside a hospital VPC or a pharma research enclave — drop-in for existing Elasticsearch clients, auditable per query, HIPAA-BAA-ready.

PHI-SOVEREIGN · HIPAA-READY · DROP-IN

02 · WHY THIS MATTERS NOW

AI IN CARE MEETS THE PHI BOUNDARY.

Every health system has an AI strategy: ambient documentation, clinical decision support, patient-portal copilots, claims automation, research RAG. Every one of them stalls at the data layer — PHI cannot leave the perimeter, and the retrieval stacks that commercial AI vendors ship require it to.

PHI SOVEREIGNTY	HIPAA · GDPR · provincial / national health-data laws. Patient data cannot be embedded in a vendor's cloud retrieval service. The retrieval layer must run inside the provider's perimeter.
CLINICAL LATENCY	Ambient scribes, CDS alerts, MAR reconciliation — retrieval latencies above ~200 ms disrupt the clinical workflow. Most vendor RAG stacks exceed that once orchestration is added.
FRAGMENTED STACKS	EHR analytics on one store, embeddings on another, audit logs on a third. A PHI-access audit crosses three systems with three RBAC models. Compliance becomes adversarial.
RESEARCH SILOS	Genomics, imaging, publication corpus, clinical trials · each in a separate retrieval tool. Researcher RAG productivity stalls on retrieval-stack migration, not on model capability.

03 · UNIFIED AI DATA STACK · PHI-SAFE

ONE PLATFORM. INSIDE THE PHI BOUNDARY.

XERJ runs inside the hospital VPC, the pharma research enclave, or the sovereign cloud of your choice. One binary hosts every retrieval primitive a clinical or life-sciences AI workload needs — and every retrieval is audit-logged as PHI-access evidence.

SOVEREIGNTY

100 %**SELF-HOSTED
PHI-SAFE**

Your VPC · your data centre · HIPAA-BAA-eligible.
Embeddings of PHI never leave the perimeter unless
a sanctioned embedder is configured.

LATENCY

< 40 MS**HYBRID P95
CLINICAL BUDGET**

Fits inside EHR round-trip budgets. Ambient-scribe
retrieval keeps up with conversation cadence.

AUDIT

PER-QUERY**EXPLAIN-PLAN
PHI-ACCESS LOG**

Every retrieval is structured audit: principal, query,
rows returned, cost. Exportable to compliance
teams.

CONSOLIDATION

4 → 1**RETRIEVAL STACKS
UNIFIED**

EHR analytics + research RAG + clinical agent
memory + SIEM on one platform.

04 · HEALTHCARE BLUEPRINT

CLINICAL BOUNDARIES. ONE PLATFORM.

Two XERJ environments split by data-classification boundary. Same binary, same audit model, different PHI scopes.

DOMAIN A · CLINICAL OPERATIONS PHI · inside hospital VPC

Sources: Epic / Cerner analytic extracts · HL7 · FHIR · device telemetry

|



XERJ cluster · 3-node HA · HIPAA BAA scope

Workloads:

- ambient scribe RAG (encounter notes + patient history)
- clinical decision support (retrieval from local KB)
- care-gap detection via longitudinal record queries
- patient-portal copilot (scoped to the patient)

Retention: encounter 7 y · audit 10 y · PHI tombstones tracked

DOMAIN B · RESEARCH & OPERATIONS de-identified · research enclave

Sources: genomics · imaging · publication corpus · trial registries

|



XERJ cluster · 3-node HA · de-identified + consented

Workloads:

- researcher RAG (pubs + internal corpus)
- trial-matching over eligibility criteria
- imaging similarity (radiologist second opinion)

Retention: research 10 y · trial lifecycle + regulatory minimums

SHARED

Enterprise KMS (BYOK roadmap Q2) · SAML/OIDC tied to directory

Per-tenant indices by service line · mutual TLS between services

ONE BINARY · TWO PERIMETERS · PHI NEVER LEAVES DOMAIN A

05 · WHERE XERJ APPLIES

CARE.
RESEARCH.
OPERATIONS.

The AI workloads that dominate health-system and life-sciences roadmaps. Each maps to one XERJ deployment pattern.

WORKLOAD	PATTERN	STORE
AMBIENT SCRIBE	Encounter transcripts + patient history retrieved in real time. Hybrid keyword + vector against chart data. Agent memory per encounter.	HYBRID
CLINICAL DECISION	Guideline + local-policy retrieval at the order-entry moment. Latency inside EHR round-trip budget. Per-specialty RBAC.	FTS
PATIENT PORTAL COPILOT	Patient-scoped agent retrieves from their own chart + approved KB. Session-isolated agent memory. Time-decay for recent encounters.	HYBRID
RESEARCHER RAG	PubMed mirror + internal corpus · hybrid retrieval for literature review. De-identified workspace. Re-embed on corpus refresh.	VECTOR
TRIAL MATCHING	Eligibility criteria as retrievable structured queries. Match patient cohort against protocol criteria in milliseconds.	HYBRID
IMAGING SIMILARITY	Radiology-image embeddings stored alongside metadata. Hybrid: keyword filter (modality / anatomy) + vector nearest neighbors.	VECTOR
CLAIMS / BILLING	Denials analysis · remark-code clustering · retrieval-augmented appeal-letter generation. Revenue-cycle automation on unified data plane.	COLUMNAR
SOC / SIEM	Health-system cyber SOC · MTTD in milliseconds · unified with cloud audit + identity logs. Inherits the HIPAA Security Rule posture.	LOGS

06 · COMPLIANCE & PRIVACY

PHI-GRADE.
BY DEFAULT.

Controls required to meet HIPAA Security Rule / Privacy Rule, HITECH breach notification, and equivalent national health-data regulations. Self-hosted means most obligations flow from customer controls — XERJ supplies the engineered-in controls auditors expect to see.

STANDARD	RELEVANCE	STATUS
HIPAA · HITECH	BAA on Enterprise tier · self-hosted inherits customer Security Rule posture	BAA 2027
PHI ENCRYPTION AT REST	AES-256-GCM on segment + WAL · keys in external KMS	BETA
BYOK	Customer-managed keys · no XERJ-side plaintext · rotation without downtime	ROAD Q2
GDPR ART. 9 HEALTH	Explicit consent · right to erasure · cross-border transfer controls · self-hosted in region	GA
SOC 2 TYPE I	Platform controls for SaaS offering	ROAD Q3 2026
ISO 27001	ISMS for platform engineering	ROAD Q4 2026
HITRUST CSF	Inherits agency controls · commercial roadmap 2027	ROAD 2027
FDA GOOD MACHINE LEARNING	Per-query explain-plan aligns with change-control + audit-trail expectations	GA
AUDIT ARTIFACT	Every retrieval produces a structured log: principal, query, index, row-count, cost. Exportable as HIPAA PHI-access audit. Replayable for breach-investigation drills.	
DATA LIFECYCLE	Per-tenant, per-index retention policies. Patient-right-of-erasure supported through tombstone + segment-level compaction. Tombstones survive rebuilds.	

07 · CLINICAL STACK · INTEGRATED

EHR.
EMPI.
ALL-IN.

The clinical and research tools a health system already runs integrate via standards (FHIR, HL7, DICOM-metadata, ES wire, OTLP) rather than bespoke connectors.

CATEGORY	TOOLS	STATUS
EHR	Epic · Cerner / Oracle Health · Meditech · Athenahealth — via FHIR bulk export + HL7 v2 feed + CDC streams	GA
EMPI / MDM	IBM Initiate · Rhapsody · Redox · via patient-ID resolver service that queries XERJ	GA
IMAGING	DICOM-metadata ingest · PACS integration via HTTP proxy · image embeddings stored with study metadata	GA
LAB / LIS	HL7 ORU messages · LOINC-coded results · ingested as structured events	GA
IDENTITY	OIDC · SAML · enterprise AD · SMART-on-FHIR for clinical apps · SCIM provisioning	ROAD Q2
AI FRAMEWORKS	LangChain · LlamaIndex · Haystack · DSPy · via ES retriever · clinical embedder options (OpenAI · Voyage · Gemini · self-hosted)	GA
DE-IDENTIFICATION	Safe-Harbor / Expert-Determination rules applied pre-index via Vector / Otel processor · plug-in de-identifier services	GA
RESEARCH	PubMed mirror · clinicaltrials.gov · internal publication store · OMOP / CDM data marts	GA
REVENUE CYCLE	837/835 claim data · denial-reason code retrieval · prior-auth workflow agents · via ES API	GA

08 · MEASURED VS LEGACY

NUMBERS CLINICIANS DON'T NOTICE.

Clinical AI succeeds when the retrieval layer is invisible. These are the latency and footprint numbers that make invisibility possible.

METRIC	XERJ	LEGACY STACK
HYBRID RETRIEVAL P95 · 1 B VECTORS	38 ms	100–250 ms (orchestrator)
AMBIENT-SCRIBE LOOKUP · IN-SESSION	< 40 ms	typically > 150 ms
VECTOR MEMORY · 10 M × 1.5 K	18 GB	92 GB — 5.1×
RECALL @ 10 VS FLOAT32	< 0.5 pp loss	baseline
COLD START	50 ms	~15 s JVM warm-up
PER-QUERY PHI AUDIT	native · explain-plan	application plumbing
PHI DATA EGRESS	0 (self-hosted)	varies by vendor
DEPLOY FOOTPRINT	13 MB binary	~620 MB JAR + JRE
CLINICAL IMPACT	Ambient scribes keep pace with conversation. CDS alerts fire inside order-entry latency budgets. Researcher RAG replaces a literature review that used to take hours.	
FINANCIAL IMPACT	Vector-DB + session-cache + SIEM contracts collapse to one. ~80 % infrastructure reduction on the consolidated stack. BAA scope reduced to one vendor.	

09 · FIVE PHASES · ONE YEAR

NON-PHI.
DE-ID.
PHI.

Health-system rollouts start outside the PHI boundary, graduate to de-identified workloads, then move into the PHI environment once the BAA and CISO-office reviews complete.

01	NON-PHI PILOT WEEKS 1-2	Dev environment · open-data corpus (PubMed subset) · reproduce RAG_BENCH · CISO + privacy-office architectural review · draft data-flow diagrams.
02	DE-ID UAT MONTHS 1-3	De-identified clinical corpus · Safe-Harbor pipeline · initial RAG workloads for researchers · BAA conversation opened · initial DPIA / privacy impact assessment.
03	PHI PILOT QUARTER 2	BAA executed · production-adjacent PHI scope · ambient-scribe or patient-portal pilot · KMS integration · audit-trail validation with privacy office.
04	CLINICAL PRODUCTION QUARTER 3	First PHI workload at production scale · EHR-adjacent deployment · per-encounter audit logs flowing into compliance dashboards.
05	PLATFORM CONSOLIDATION QUARTER 4	Second + third PHI workload · vector-DB contract retired · unified retrieval platform owned by the data team · board metrics live.

PHI NEVER LEAVES THE PERIMETER · REVERSIBLE AT EVERY PHASE

10 · NEXT STEPS

PROVE IT. OUTSIDE PHI.

A two-week non-PHI pilot. Open-data corpus. Architecture review with your CISO and privacy office. A dated BAA conversation on day one if Enterprise tier is in scope.

HEALTHCARE CONTACT

HELLO@XERJ.AI

XERJ.AI · XERJ.COM · XERJ.DEV

WHAT YOU GET

Private binary · PubMed-subset embedding pipeline · Vector de-id processor config · HIPAA Security-Rule control crosswalk · audit-log export example.

COMPANION
BRIEFS

xerj-tech-brief.pdf (architect) · xerj-usecase-rag.pdf (RAG) · xerj-usecase-security.pdf (cyber SOC)