LlamaFarm By Rownd, Inc

Zero-Trust AI Projects for Healthcare: Run AI inside your environment (private cloud/on-prem), **keep PHI in your runtime**, and deploy the **same bundle** across data centers, clinics, and air-gapped/isolated networks.

THE CHALLENGE

- **Regulatory & PHI constraints:** HIPAA/HITECH, 42 CFR Part 2, BAAs, and privacy rules make cloud-SaaS AI risky; AI must **run inside** your boundary with **no PHI egress**.
- **Integration complexity:** Real-world AI spans **EHR (FHIR/HL7v2)**, **PACS/DICOM**, labs, bedside devices, and ITSM—gluing RAG, embeddings, **fine-tuning**, and runtime is brittle.
- **Data silos:** Notes, imaging, vitals, and call transcripts live in separate systems; opportunities are missed without cross-source correlation. Integration directly with EHR to use real information.
- **Downtime reality:** Rural clinics, disaster events, and security incidents require **offline-first** AI with **identical outputs** during EHR downtime procedures.

THE LLAMAFARM SOLUTION

Customer-Hosted with PHI Residency: LlamaFarm runs entirely inside your security boundary (on-prem or private cloud). No external calls. PHI, logs, and models remain in your tenant; BAAs supported.

Clinical AI-as-Code (YAML): Declare sources (FHIR/DICOM/HL7), guardrails (PHI redaction, role-based access), and prompts in YAML. Everything is auditable, versionable, and easy to roll back.

True Multimodal Intelligence: Deploy specialized and fine-tuned LLMs that are experts in focused areas using your real data. Add voice and vision on top to create truly multi-modal AI projects that can identify threats, talk to users, and be extendable to fit any use-case.

Bundle Once, Deploy Anywhere: Combine EHR notes (FHIR), imaging (DICOM), labs (HL7), audio (clinician–patient), and policy docs into one reasoning stack; fine-tune models to become service-line experts.

TECHNICAL CAPABILITIES

- **EHR**: FHIR R4 (read-only to start), HL7v2 (ADT/ORU/ORM), CCD/CCDA, ingestion
- **Docs/RAG:** OCR, deidentification, vector search, lineage & citation
- **Audio/Voice**: Encounter transcription, speaker labeling; voice commands (4Q2025)
- **Models**: Local/hosted LLMs, fine-tuning adapters for specialty domains
- **Agent Ready**: Agentic workflows without external calls
- **Docker Runtime**: API first, secure, config driven runtime

MISSION-CRITICAL USE CASES

🩺 **Clinical Intelligence & Care Coordination:** Unify EHR notes, imaging reports, labs, and call transcripts to surface key entities (problems, meds, allergies) and draft handoffs/letters with citations.

Voice • Image Recognition • Automated translation • Pattern detection • Report •

🛠️ **Biomedical Asset Uptime Copilot:** A voice-driven copilot for biomed/HTM teams that “reads” service manuals, parses logs, and recommends fixes; learns from prior work orders and predicts failure windows.

Voice-to-text: RAG Documents • Vector Database • Finetuned Model • Failure prediction • Offline operation

📚 **Clinical Best-Practices & Quality Copilot:** RAG over order sets, pathways, AHRQ/IDSA/house guidelines with local policy overlays. Generate discharge instructions, education, and handoff summaries with citations.

Multi-stream synchronization • Automated transcription • Document RAG • Prompt Engineering

🧾 **Revenue Cycle & Compliance Assistant:** Drafts prior-auth packets, highlights documentation gaps for CDI/coding, and audits PHI exposure paths.

Document Creation • Agentic Flows • Document RAG • Prompt Engineering • Fine-tuned models

LlamaFarmAI Appliance (Optional)

Turnkey on-prem GPU appliance, pre-installed for multimodal workloads; integrates with hospital K8s/VMware; at-rest encryption with FIPS-validated modules.

WHY CHOOSE LLAMAFARM?

- **Immediate Value**: Stand up pilots in days, not months; no AI team required.
- **Complete Control**: Your data, your models, your infrastructure.
- **Governance & Safety**: PHI tagging/redaction, RBAC/ABAC, full audit to SIEM.
- **Open-core**: LlamaFarm is open-source, can extend and customize, no vendor lock-in
- **Cost-Effective**: Start small (single service line), scale to enterprise

Pricing: Open-core; per-site/per-node runtime + enterprise support, with optional pro modules - volume tiers and ELA available.

Getting started: BAA + a 60–90-day in-tenant pilot (EHR/FHIR). Either test or real PHI. All data stays in your environment; exit with

Website: https://Llamafarm.dev/enterprise

Open Source: https://github.com/llama-farm/llamafarm