MedGemma FHIR-Bridge: Turning Medical "Dark Data" Into Lifesaving Intelligence

MedGemma Impact Challenge 2025 (Kaggle) | Submission by Team **MedGemma FHIR-Bridge**

Goal: convert unstructured clinical artifacts (scanned PDFs, handwritten notes, images) into **safe, interoperable HL7 FHIR R4 bundles** - without letting AI hallucinations pollute patient records.

The Crisis: Healthcare's "Dark Data" Problem

Across hospitals and clinics, critical medical facts live in analog or semi-digital artifacts: prescriptions, lab printouts, radiology impressions, and clinician notes. Most of this information is unstructured, hard to search, and difficult to exchange between systems.

- **Manual transcription burden:** staff often re-type labs into EHRs, consuming minutes per report and creating avoidable errors.
- **Missed or delayed alerts:** a single misread value (e.g., decimal shifts or missing zeros) can change triage decisions.
- **Interoperability gridlock:** without standards like **HL7 FHIR**, data stays trapped in silos across sites and vendors.

The AI Promise - and Peril

Multimodal models (e.g., MedGemma 1.5) can read messy documents, but healthcare cannot tolerate hallucinations. We need extraction that **verifies, repairs, and fails safely**.

Our Solution: a "Self-Healing" Clinical Data Pipeline

Layer 1	Layer 2	Layer 3
Reader	Auditor	Collaborator
MedGemma 1.5 via vLLM	Semantic firewall + FHIR validation	Human notes + AI synthesis

We treat the model as a "junior resident": it drafts structured facts, then deterministic logic validates them before anything reaches the patient record.

Output: schema-valid, interoperable **FHIR R4 bundles** (Observations, MedicationRequests, DiagnosticReports) ready for downstream EHR ingestion.

Verify & Repair Architecture

Layer 1 - Reader (MedGemma 1.5 via vLLM)

- **Multi-image collective reasoning:** up to 8 related images per case so the model can cross-check evidence across pages.
- **Modality-aware prompting:** classify document type (Lab/Radiology/Prescription/Vitals), then apply a specialized extraction protocol.

Layer 2 - Auditor (Semantic Firewall)

- **Hybrid parser:** accepts JSON / TSV / imperfect Markdown and normalizes to structured JSON.
- **Medical axiom checks:** context rules catch common clinical OCR failures (unit mismatches, missing scientific notation, impossible ranges).
- **FHIR schema enforcement:** map to LOINC / SNOMED-ready concepts and validate required fields (e.g., Observation.valueQuantity).
- **Green-signal fallback:** if extraction fails, emit a minimal but valid bundle with evidence metadata - preserving workflow without corrupting data.

Layer 3 - Collaborator (Human-AI Synthesis)

- **Clinician notes** attach context directly to the record (symptoms, history, judgement calls).
- **AI synthesis** re-reads the original artifact plus clinician notes to generate a concise clinical summary and flagged uncertainties.

Why this beats "black-box" extraction

Challenge	Traditional OCR/AI	MedGemma FHIR-Bridge
Hallucination risk	Little/no validation; errors can persist	Multi-stage semantic firewall + schema enforcement
Unit ambiguity	Reads values literally (e.g., "370 /uL")	Context-aware scaling & axiom checks (e.g., platelet 370,000 /ul)
Complex lab tables	Often loses nested rows/columns	Strict TSV extraction preserves all rows for CBC differentials
Interoperability	Proprietary JSON; siloed	HL7 FHIR R4 bundles (LOINC/SNOMED-ready)
Privacy/edge	Cloud GPU common; higher PII risk	On-prem vLLM; no data leaves hospital network

Two key engineering moves

1) TSV Protocol for dense lab tables

For Complete Blood Count (CBC) and similar reports, we request **strict TSV** (TEST | VALUE | UNIT | RANGE). This avoids JSON structural hallucinations and preserves every row for downstream validation.

2) Smart Rerun (temporal correction)

Clinicians can reprocess historical records with improved prompts and rules. The system re-validates the updated bundle and refreshes review timestamps so improved results surface immediately.

Measured Impact

- **Time saved:** complex lab report entry shrinks from minutes to seconds (AI extraction + validation). At scale, this recovers substantial clinical hours.
- **Error reduction:** firewall rules catch frequent failures: decimal shifts, unit scaling, and reference-range mismatches.
- **Interoperability unlocked:** FHIR R4 bundles enable API-first exchange and integration with modern EHR ecosystems.
- **Edge-native privacy:** on-prem vLLM deployment minimizes cloud exposure and supports GDPR/HIPAA-aligned operation.

Validation & Demonstration Plan

We validate on noisy, real-world document conditions across prescriptions, radiology, and lab tables using public Kaggle datasets.

Dataset category	What it stresses	What we measure
Handwritten prescriptions	Illegible text, incomplete dosage	Medication + dosage extraction accuracy; uncertainty flags
Radiology artifacts	Free-text findings, severity cues	Structured findings + concise clinical summary
Lab reports (CBC/biochem)	Nested tables, multi-page continuity	Row-complete TSV + schema-valid FHIR Observations

Radical transparency: the Evidence Tab

Reviewers can toggle between (1) clinical view, (2) raw TSV extracted by the model, and (3) validated FHIR JSON. This makes non-hallucination auditable instead of assumed.

Why this submission should win

- **Safety-first engineering:** deterministic validation + fail-safe bundle generation.
- **Production-ready stack:** modular API service, CI/CD, and deployment via containers.
- **Workflow-native:** batch uploads, collaborative notes, and smart reruns for continuous improvement.
- **Open-standards obsession:** LOINC/SNOMED-ready concepts and FHIR R4 outputs to prevent new silos.

Submission links:

Repository: [GitHubURL]

Demo video: [YouTubeURL]

Contact: [Gowtham Sridhar – gowtham.sridher5@gmail.com]

Linkedin – https://www.linkedin.com/in/gowtham-sridher

MedGemma FHIR-Bridge turns multimodal reading into clinical infrastructure: interoperable, auditable, and built to fail safely.