Radiology Value Pipeline

A local-first, auditable radiology workflow platform for chest X-ray and brain MRI

AUTHOR
Saule

PUBLISHED
July 25, 2026

Radiology Value Pipeline is a production-style MVP for a radiology workflow copilot. It is not a toy classifier and not an autonomous diagnostic product. It is an end-to-end platform skeleton that ingests imaging studies, runs modality-aware model pipelines, generates structured assistive outputs, captures human feedback, preserves audit and lineage, and publishes dashboard-ready operational, clinical quality, and governance data.

The current build supports **chest X-ray** and **brain MRI** in the same workflow. Chest X-ray demonstrates high-throughput radiology operations. Brain MRI demonstrates advanced imaging AI, multi-class tumor phenotype classification, segmentation-backed localization, and a patient-level 3D corroboration path.

Important

This project is built around a strict clinical boundary: AI outputs are assistive suggestions only. There is no autonomous diagnosis, no auto-signing, and no unsupported clinical performance claim. The work is a technical platform demonstration, not a cleared medical device.

1 Executive Summary

Radiology Value Pipeline is my attempt to answer a practical question: what would a serious, human-centered radiology AI platform look like if it were designed around value, auditability, and clinician trust from day one?

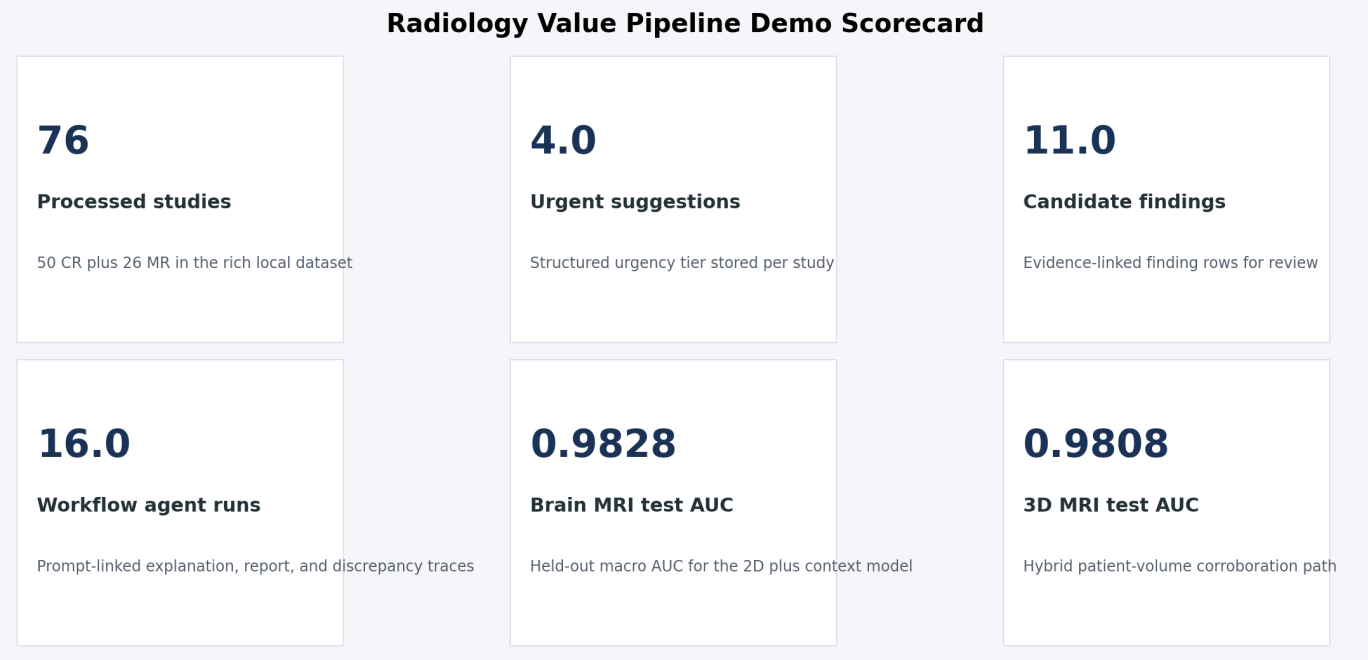
The answer is not a single model. It is a pipeline.

The system is a local-first modular monolith with a FastAPI backend, PostgreSQL operational database, DB-backed worker orchestration, browser review workspace, Power BI reporting package, model lineage, prompt lineage, governance snapshots, audit events, and feedback loops. The product is built to make the AI pipeline visible rather than magical.

The result is a credible foundation for a radiology workflow platform that can:

- Register chest X-ray and brain MRI studies.
- Preserve normalized metadata, retrieval references, artifacts, and lineage.
- Run ordered specialist model services and workflow agents.
- Store structured findings, urgency suggestions, report drafts, discrepancies, feedback, model runs, and audit records.
- Surface study-level review in a polished browser UI.
- Publish Power BI-ready exports and dashboards for operational, clinical quality, and governance reporting.
- Maintain a real-vs-stub boundary so the platform can evolve without pretending that all pieces are clinically validated.

76 processed demo studies	**50** chest X-ray studies	**26** brain MRI studies
129 candidate findings persisted	**858** dashboard fact rows	**304** workflow agent runs

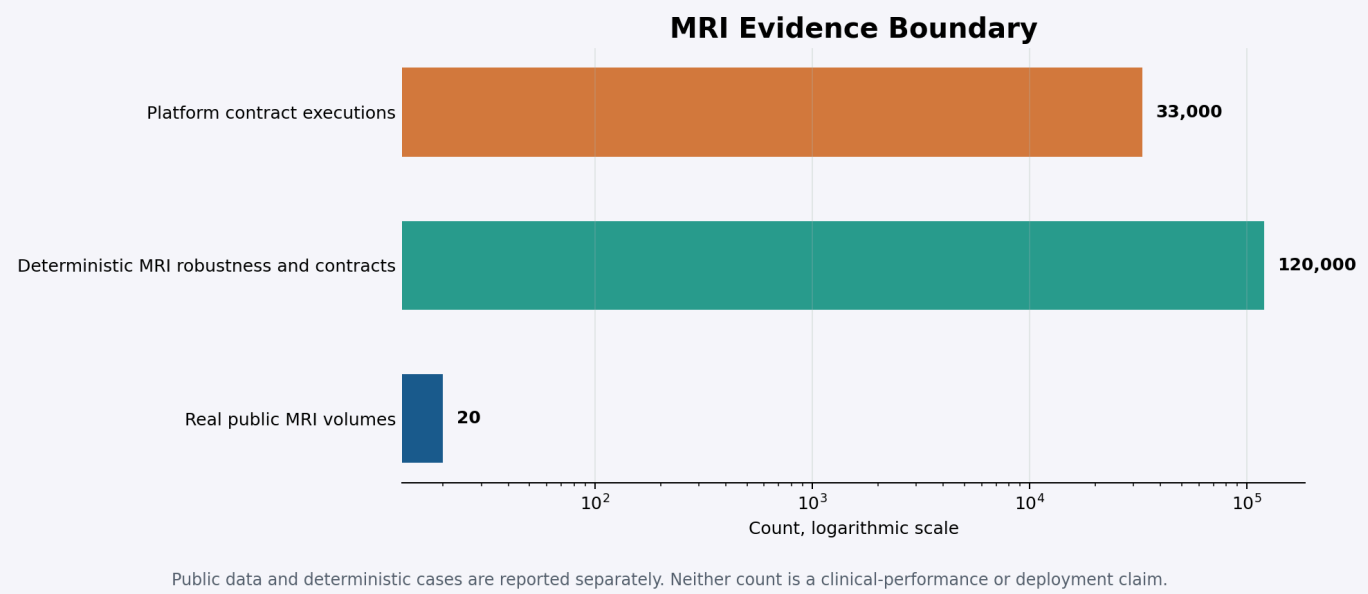


2 MRI Evidence Boundary

The MRI work has two distinct evidence categories. They are presented together for transparency, but they are not interchangeable and must not be described as a single patient cohort.

Evidence category	Count	What it establishes	What it does not establish
Real public MRI volumes	20	GPU execution of the downloaded MONAI/NVIDIA BraTS bundle on public MSD Task01 BrainTumour volumes, including prediction conversion, scoring, review overlays, and candidate cards	Clinical performance, prospective utility, or deployment readiness
Deterministic MRI robustness, acceptance, conversion, and contract evaluations	120,000	Reproducible software behavior across synthetic failure modes, workflow contracts, and acceptance gates	120,000 real people, scans, or clinical outcomes
Total engineering case-evaluations	123,020	Breadth of the tracked validation and contract artifacts	A clinical validation cohort

The completed public cohort ran on CUDA with the bundle ROI `240 x 240 x 160`. All 20 cases completed, with mean whole-tumor Dice `0.9202`, mean region macro Dice `0.8411`, mean whole-tumor HD95 `7.3455 mm`, and mean whole-tumor 2 mm surface Dice `0.2970`. Every candidate card remains `reject_for_more_evidence`. That is intentional: the system retains the artifacts and the result without claiming it is ready for clinical promotion.



3 Why This Project Matters

This project is personal.

My father and his brother both died from cancer in their mid 40s, and poor or delayed diagnostic pathways were part of the story. I do not say that to claim a model would have saved them. I say it because diagnostic systems are human systems, and human systems need better memory, better feedback loops, better measurement, and better ways to reduce preventable misses.

Radiology is one of the places where this matters most. A radiologist is asked to move fast, read complex imaging, integrate clinical context, handle prior exams, manage interruptions, support urgent findings, and produce reports that downstream teams act on. The right AI system should not try to replace that expertise. It should make the workflow more inspectable, more measurable, and more resilient.

That is the motivation behind Radiology Value Pipeline:

- Help clinicians see candidate findings faster.
- Preserve uncertainty instead of hiding it.
- Track what the AI suggested, what the clinician accepted or rejected, and what changed later.
- Build governance into the workflow instead of bolting it on after deployment.
- Make performance visible by modality, site, subgroup, outcome, and time.

This is the kind of platform I would want to build around a real clinical team: technically ambitious, clinically humble, and designed to improve through feedback.

4 Product Snapshot

Radiology Value Pipeline currently includes:

- A FastAPI backend with versioned APIs for studies, pipeline execution, findings, report drafts, feedback, final reports, discrepancy analysis, dashboards, and governance.
- A normalized PostgreSQL schema with UUID primary keys, JSONB raw payloads, append-only events, model and prompt version foreign keys, RLS-ready site scoping, and partition-ready append-heavy tables.
- A DB-backed worker pattern using row locking rather than a separate queue broker.
- A dual-mode ingestion architecture with a demo adapter and a DICOMweb-oriented Orthanc adapter contract.
- A chest X-ray model path based on ChestMNIST compact CNN ensembles.
- A brain MRI path based on the public Figshare Brain Tumor MRI benchmark, including context-aware slice classification, segmentation-backed localization, and patient-level 3D corroboration.

- A review UI that shows imaging preview, AI synthesis, model posture, target-level probabilities, draft report text, feedback controls, discrepancy events, audit trail, and workflow agent ledger.
- Power BI dashboard exports and a PBIP handoff package with operational, clinical quality, governance, model inventory, and agent lineage data.
- Documentation anchors including `agents.md`, `specifications.md`, architecture notes, ADRs, schema explainer, dashboard notes, API reference, and final validation notes.

RVP

Radiology Value Pipeline
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

Human review remains required

Assistive output only

Chest X-ray + Brain MRI

Clinical operations workspace

Brain MRI T1 Post-Contrast • processed • MR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**MR-1B88A9E1**	**baseline**	**2:55:52 PM**

Radiology worklist

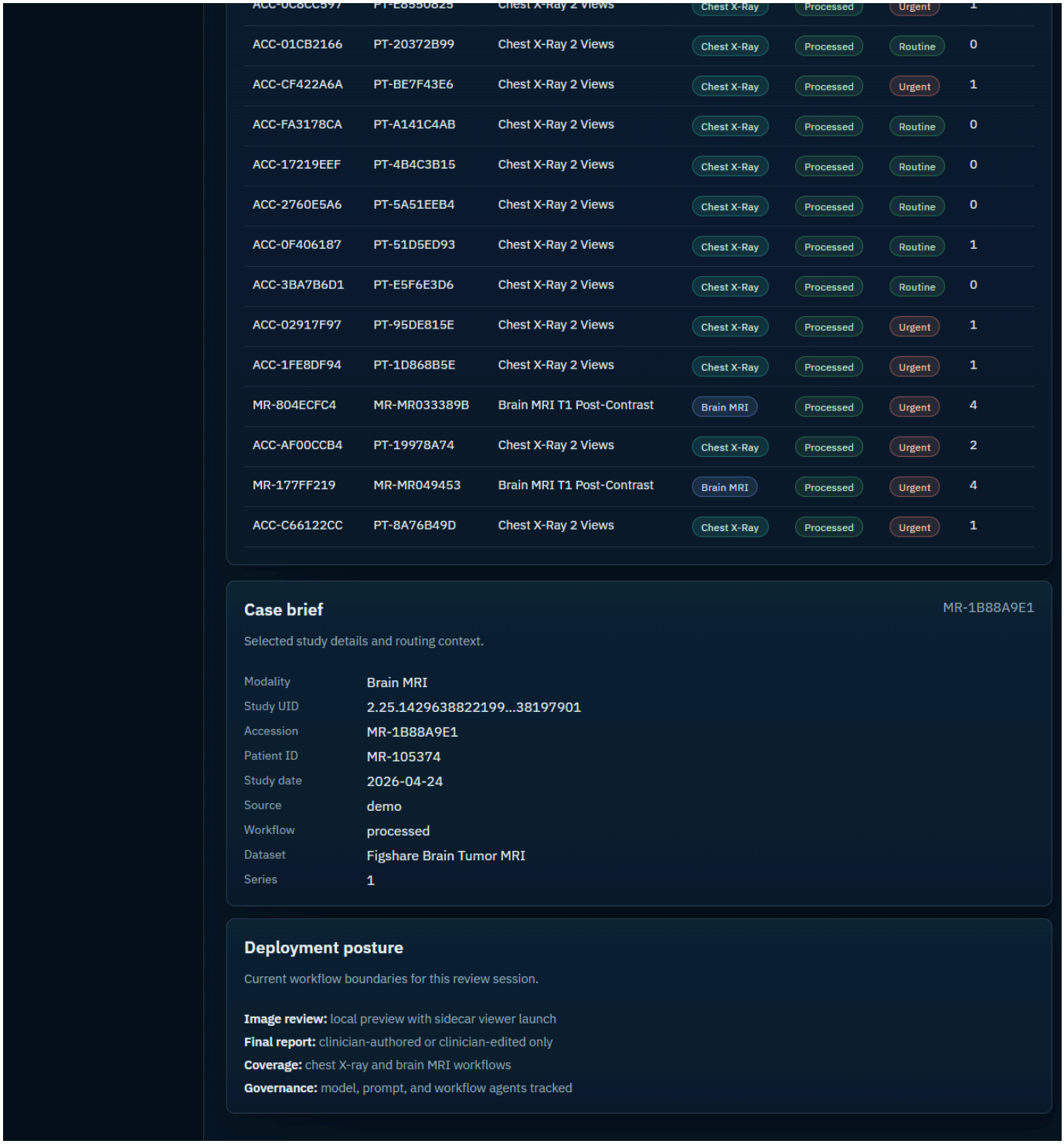
76 studies

Operational queue, modality mix, urgency, and study selection.

Open queue	Urgent	Chest	MRI	Final reports
0	**30**	**50**	**26**	**76**
Studies pending or active	Flagged for expedited review	Chest X-ray studies in queue	Brain MRI studies in queue	Signed study count

Accession	Patient	Description	Modality	Status	Urgency	Findings
MR-1B88A9E1	MR-105374	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-6A3C83FD	MR-103257	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-316055C7	MR-MR017260F	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-C2DE7B07	MR-105936	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-2D5DEB13	MR-MR033389B	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-7B20AC5F	MR-MR033389B	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-58096482	MR-MR017260F	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-9ACD29C3	MR-104281	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-E540B5FF	MR-MR037458C	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-9EDD20FE	MR-105936	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-F0396669	MR-111532	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-1A832D4D	MR-111532	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-B62BBDE2	MR-107747	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-84528A3B	MR-MR049453	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-4558DC91	MR-MR049453	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-20F67CE6	MR-103046	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-F046A07F	MR-111532	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-46CB4A8B	MR-MR033389B	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Urgent	4
MR-4617B919	MR-109471	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-06AA0F1C	MR-102675	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-2A4B4E38	MR-104281	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-BFB74CF4	MR-103257	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-12C4E916	MR-99017	Brain MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
MR-4ACEE592	MR-101145	Sellar MRI T1 Post-Contrast	Brain MRI	Processed	Routine	4
ACC-661D5B5B	PT-3735DE05	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1

ACC-A60F7789	PT-F2306D59	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-BECE98E4	PT-403AC06C	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-A842026C	PT-1F2569A0	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-6607FA80	PT-79060E2B	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-1B2D22DD	PT-374C838C	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-F1F59E18	PT-C7EA8494	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-662DCB8B	PT-8CEB4CF0	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	1
ACC-F53936FC	PT-8AE155E7	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-C17059B7	PT-8EBF8C13	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-0EDEC24A	PT-40790F46	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-77A2589C	PT-49849297	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-584133C3	PT-56DBAC67	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-22F493F3	PT-F1D7EAEA	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-AAEE82C4	PT-D559E66C	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-93FE37FB	PT-1489B5FB	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-12971E97	PT-150CB8E8	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-7E284394	PT-CB7DED42	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-312BACA1	PT-7DF8029D	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	2
ACC-3105E5B8	PT-B29FA24E	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-2E6CA905	PT-AAB46774	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-263C7964	PT-2553BB82	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-5C01C0E5	PT-19367E08	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-B962017D	PT-36FDA3FE	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-59E6F592	PT-BD9C2436	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-6CE35AED	PT-222F2EF4	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-C0BB5B58	PT-21D8C927	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-D1DFAB94	PT-27F1CEF5	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-1FD5A8CB	PT-69A39A2B	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-511A393C	PT-8565A34B	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-44659B19	PT-E35DF65F	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-AA2FCA30	PT-7B782C5E	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-1857591E	PT-6E7A1033	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	2
ACC-BD575C37	PT-B16F4F67	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-F92C7D11	PT-4CB9B2C7	Chest X-Ray 2 Views	Chest X-Ray	Processed	Urgent	1
ACC-9A228201	PT-1681136B	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-BF0D87A6	PT-331F7151	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0
ACC-0315416F	PT-D098BA18	Chest X-Ray 2 Views	Chest X-Ray	Processed	Routine	0



5 The Design Choice That Defines The Project

The central architectural decision is that this is not a model demo. It is a workflow platform.

In healthcare AI, the hard parts are not only model architecture. The hard parts are versioning, traceability, feedback, governance, review experience, data contracts, operational monitoring, and making it possible for a clinician to understand what happened after the fact.

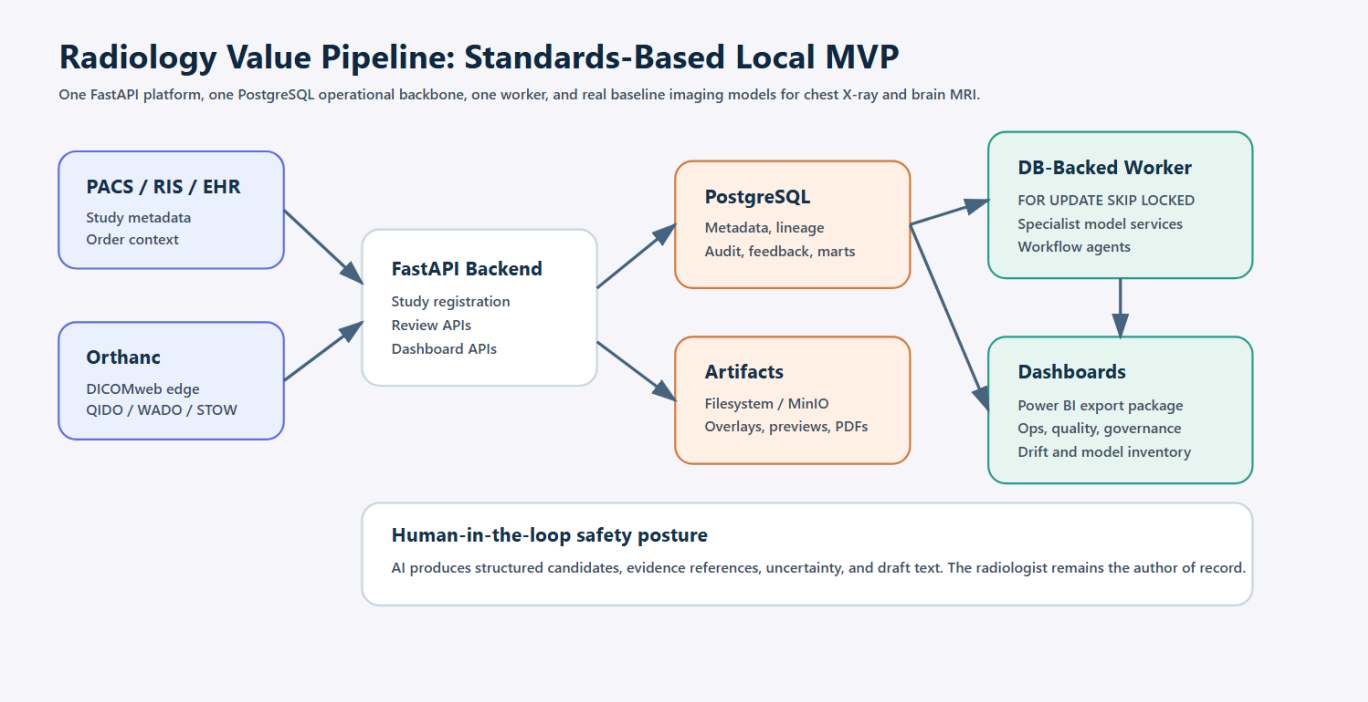
This platform treats every AI output as part of a governed workflow:

- The model run is stored.
- The model version is stored.
- The prompt version is stored when applicable.
- The parameters and raw output payload are stored.
- The structured output is stored separately from human-readable text.
- The clinician feedback is append-only.
- The final report remains clinician-authored or clinician-edited.
- The discrepancy analysis links AI output to the final report and downstream quality review.
- The dashboard marts expose adoption, override, drift, discrepancy, latency, and governance signals.

That design is what makes the project more than a classifier.

6 Architecture

The MVP is intentionally a modular monolith plus worker. It avoids Kafka, event-sourcing theater, and microservice sprawl. Clean boundaries exist inside the repo, but the system stays comprehensible and runnable on a laptop.



6.1 Runtime Architecture

```
Browser review workspace
        |
        v
FastAPI backend
        |
        +-- Study registry and review APIs
```

```
        +-- Pipeline orchestration APIs
        +-- Feedback and discrepancy APIs
        +-- Dashboard and governance APIs
        |
        v
PostgreSQL operational backbone
        |
        +-- normalized study metadata
        +-- model and prompt versions
        +-- pipeline jobs
        +-- model runs
        +-- findings and urgency scores
        +-- report drafts and final reports
        +-- feedback and discrepancy events
        +-- audit and governance snapshots
        +-- dashboard marts
        |
        v
Worker process
        |
        +-- representation service
        +-- view quality service
        +-- abnormality detection service
        +-- urgency scoring service
        +-- prior context service
        +-- explanation service
        +-- report draft service
        +-- discrepancy service
        +-- workflow agents
```

6.2 Why This Architecture Is Strong

This architecture is deliberately boring where reliability matters and advanced where imaging AI matters.

- FastAPI keeps the API layer typed, inspectable, and easy to document.
- PostgreSQL is the source of truth for operational state, not just an afterthought.
- The worker uses database row locking, which is enough for an MVP and avoids an unnecessary broker.
- Orthanc and OHIF are kept as standards-based DICOM and viewer integration points.
- MinIO or filesystem artifacts keep binary storage outside Postgres.
- MLflow and model version tables preserve model lineage.
- Prompt version tables preserve report-assist and agent lineage.
- Evidently-style drift snapshots create a path for scheduled monitoring.
- Power BI exports make the governance and operations story visible to leaders, not just engineers.

6.3 Architectural Boundaries

Area	Responsibility
`backend`	API routes, request validation, auth context, HTML review workspace
`services/ingestion`	Study registration, DICOM gateway abstraction, normalized envelopes
`services/pipeline`	Specialist model orchestration and output persistence
`services/reporting`	Draft rendering and discrepancy comparison
`services/feedback`	Append-only clinician feedback events
`services/governance`	Model, prompt, agent, drift, and governance surfaces
`services/dashboards`	Dashboard mart refresh and summary APIs
`workers`	DB-backed queue polling and background execution
`monitoring`	MLflow, Evidently, and lineage helpers
`dashboards`	SQL marts, Power BI package, and export scripts

7 Technology Stack

This stack was chosen for practical clinical product engineering, not novelty.

Layer	Choice	Why it matters
API	FastAPI, Pydantic	Typed contracts, OpenAPI docs, predictable backend development
Database	PostgreSQL	Durable operational source of truth with JSONB, partitioning, RLS readiness, logical replication path
ORM and migrations	SQLAlchemy, Alembic	Explicit schema evolution and testable persistence
Imaging edge	Orthanc, DICOMweb	Standards-based DICOM edge rather than custom imaging APIs
Viewer	OHIF sidecar scaffold	Proven viewer integration path without premature custom

Layer	Choice	Why it matters
		extension work
Artifact storage	MinIO or filesystem abstraction	Keeps binaries outside Postgres and preserves local-first operation
Model lineage	MLflow	Registry, run metadata, artifact lineage, model tags
Drift	Evidently-style snapshots	Reference-vs-current monitoring model for input and output changes
Dashboards	SQL marts, CSV exports, Power BI PBIP	Executive and governance reporting path
Frontend	Server-rendered browser workspace with custom CSS and JS	Fast to iterate, easy to run locally, avoids overbuilding a separate frontend too early
Testing	pytest, smoke scripts	Contract and business-logic verification

> **Note**
>
> The stack is intentionally local-first. It can run with Docker Compose, but it also includes an embedded local PostgreSQL fallback so the product can still be demonstrated on a machine where Docker is unavailable.

8 Research-Backed Integration Choices

The project documentation captures current design-relevant research in `docs/architecture/research-notes.md`. The most important choices are:

- Use Orthanc DICOMweb and OHIF DICOMweb data source patterns rather than inventing proprietary imaging APIs.
- Keep Orthanc as the imaging edge and store normalized metadata plus retrieval references in the backend.
- Use MONAI Bundle concepts as the swappable model packaging contract.
- Keep MONAI Deploy packaging optional because packaging and runtime assumptions are more Linux and NVIDIA oriented.
- Treat MONAI Label as a future active-learning integration, not the runtime core for v1.
- Use PostgreSQL JSONB, RLS-ready site scoping, partitioning for append-heavy event tables, and logical replication readiness for future BI replicas.

- Use pgvector-ready columns for embeddings, while deferring ANN indexes until volume requires them.
- Prefer MLflow model aliases and tags over stage-centric workflows.
- Use prompt registry concepts for report-assist and agent lineage.
- Use Evidently-style reference-vs-current snapshots for drift.
- Use Power BI Import for historical marts and narrow DirectQuery only when operational freshness matters.
- Treat FDA AI-enabled device software lifecycle guidance and ONC HTI-1 predictive DSI transparency expectations as governance design inputs, not compliance claims.

9 Repository Structure

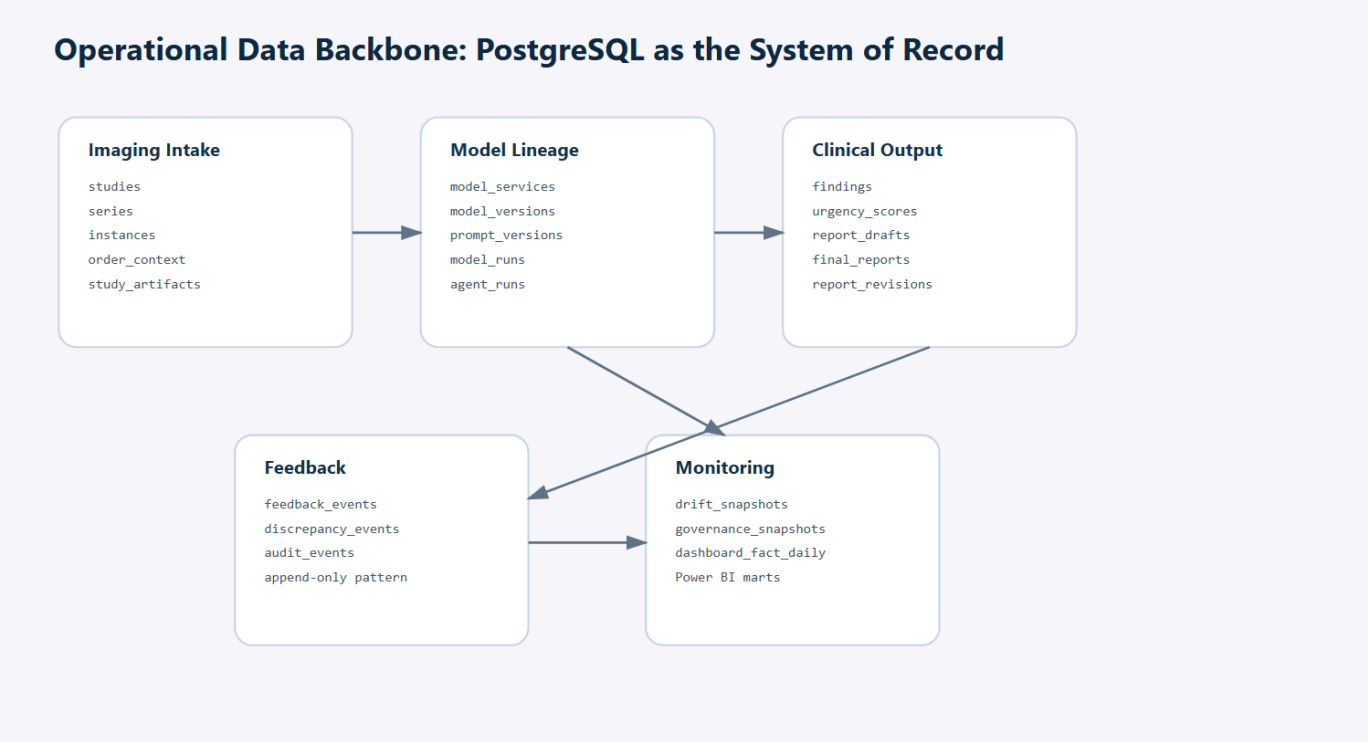
```
radiology-value-pipeline/
  backend/
    app/
      api/
      core/
      templates/
      static/
  db/
    migrations/
    models.py
    schema-explainer.md
  services/
    ingestion/
    pipeline/
    reporting/
    feedback/
    governance/
    dashboards/
  models/
    stub/
    baseline/
    monai_bundle/
  workers/
  monitoring/
  dashboards/
    sql/
    powerbi/
  docs/
    architecture/
    adrs/
  infra/
  scripts/
  tests/
  agents.md
```

```
specifications.md
README.md
```

The structure is designed to be legible to a reviewer. A hiring manager or technical interviewer can open the repository and immediately find the product specification, architecture decisions, database model, pipeline interfaces, worker behavior, dashboards, and tests.

10 Data Model

PostgreSQL is the operational backbone. The platform does not store DICOM binaries in the relational database. It stores metadata, lineage, structured outputs, and retrieval references, while image artifacts, overlays, saliency maps, masks, and generated files live in artifact storage.



10.1 Core Tables

Table	Purpose
`studies`	Study-level identity, modality, accession, status, site, and metadata
`series`	Series-level normalized imaging metadata
`instances`	Instance-level SOP references and retrieval locators
`order_context`	Indication, reason for exam, ordering context

Table	Purpose
`study_artifacts`	Preview images, overlays, masks, saliency maps, exported artifacts
`pipeline_jobs`	DB-backed worker queue and job state
`model_services`	Registry of specialist services
`model_versions`	Model bundle, artifact, configuration, validation metadata
`prompt_versions`	Prompt lineage for report assist and workflow agents
`model_runs`	Per-service run metadata, latency, inputs, outputs, and errors
`findings`	Structured candidate findings with confidence and evidence refs
`urgency_scores`	Urgency suggestions and rationale
`report_drafts`	AI-generated draft findings and impression text
`final_reports`	Clinician-authored or clinician-edited final report text
`report_revisions`	Report change history
`discrepancy_events`	AI-vs-final and AI-vs-outcome discrepancy records
`feedback_events`	Append-only clinician feedback
`drift_snapshots`	Reference-vs-current monitoring records
`governance_snapshots`	Model and prompt governance records
`audit_events`	Security and workflow audit trail
`dashboard_fact_daily`	Curated daily dashboard fact table
`agent_runs`	Workflow agent execution lineage
`agent_versions`	Agent identity, prompt linkage, and active version metadata

10.2 Schema Principles

- UUID primary keys where durable identity matters.

- `created_at` and `updated_at` timestamps on operational entities.
- `event_date` on append-heavy tables for partitioning and dashboarding.
- `site_id` and role-scope columns where future RLS policies need them.
- Natural-key uniqueness on DICOM identifiers.
- JSONB for raw model payloads, drift details, prompt render context, and governance metadata.
- Vector-ready columns for study and report embeddings.
- Foreign keys from outputs to model versions and prompt versions.
- Append-only feedback and audit records.

10.3 Example: DB-Backed Worker Claim

The worker is intentionally simple. It claims jobs with row-level locking and `SKIP LOCKED`, which gives safe concurrent polling without adding Redis or Kafka.

```
select *
from pipeline_jobs
where status = 'queued'
order by created_at asc
for update skip locked
limit 1;
```

This is an MVP-appropriate choice because the queue semantics are directly tied to the operational database, the failure modes are easy to inspect, and the system stays small enough to understand.

10.4 Example: Dashboard Mart Shape

The dashboard marts are SQL-first. Power BI is a presentation layer, not a hidden source of truth.

```
select
  event_date,
  site_code,
  studies_registered,
  studies_processed,
  urgent_cases,
  failed_runs,
  latency_p50_ms,
  latency_p90_ms
from dashboard_ops_daily
order by event_date desc, site_code;
```

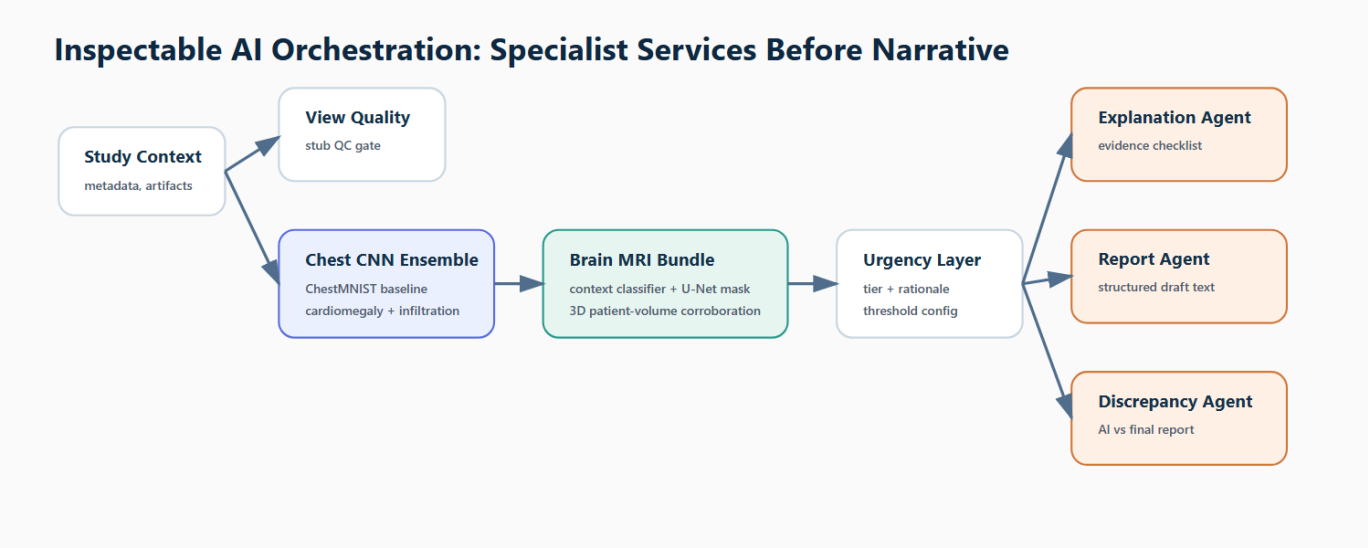
The exported SQL files are included with this write-up:

- `assets/sql/001_ops_dashboard.sql`
- `assets/sql/002_quality_dashboard.sql`

- `assets/sql/003_governance_dashboard.sql`

11 Model System

The model system is orchestrated as specialist components, not a giant monolith. Each service has a typed contract, persists run metadata, records latency and status, and can run in stub or real mode.



11.1 Specialist Services

Service	Role
`representation_service`	Creates or references imaging representation metadata
`view_quality_service`	Checks view and image quality assumptions
`abnormality_detection_service`	Produces target-level abnormality probabilities
`urgency_scoring_service`	Converts evidence into assistive urgency suggestions
`prior_context_service`	Attaches prior-context state
`explanation_service`	Produces saliency, evidence references, or explanation payloads
`report_draft_service`	Renders structured findings and concise draft text
`discrepancy_service`	Compares AI outputs with final report text

11.2 Workflow Agents

The platform also includes explicit workflow agents. These are not free-floating chatbots. They are governed workflow components with prompt versions, agent versions, input payloads, output payloads, status, latency, trace JSON, and audit linkage.

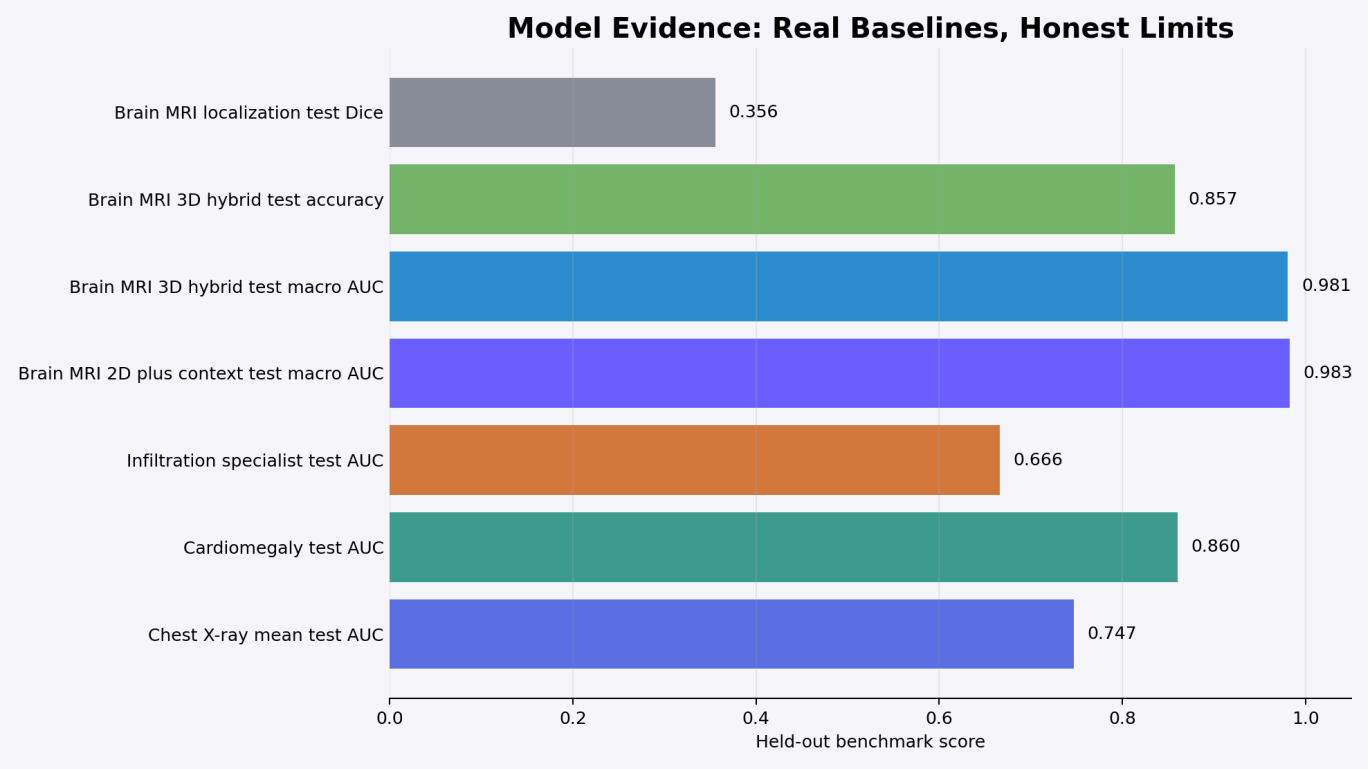
Current agent examples:

- `explanation_review_agent`
- `report_assist_agent`
- `discrepancy_review_agent`

The agent layer matters because modern clinical AI products will increasingly combine image models, structured rules, report drafting, and natural-language review. The platform makes those components observable rather than invisible.

12 Model Performance

The project uses real public benchmark data where possible, while keeping strict language around clinical validation. The AUC values below are benchmark and demo evaluation metrics. They are not claims of clinical readiness.



12.1 Chest X-Ray Baseline

Chest X-ray abnormality baseline

Runtime bundle: compact-cnn-v4-ensemble-2

Dataset: ChestMNIST benchmark

Validation mean AUC: 0.7597

Held-out test mean AUC: 0.7470

Why it matters: this is a lightweight CPU-friendly baseline that can run locally and still produce real target-level probabilities, thresholds, and governance metadata.

Chest X-ray is the operational MVP. It demonstrates the high-throughput workflow: study registration, model execution, findings, urgency suggestion, draft report support, feedback capture, discrepancy analysis, and dashboard publication.

Notable target results:

Target	Validation AUC	Test AUC	Notes
Cardiomegaly	0.8721	0.8601	Stronger chest target in the current baseline
Infiltration baseline	0.6473	0.6338	Harder target with lower performance
Infiltration specialist	0.6553	0.6658	Specialist model improves the infiltration slice

The system does not hide weak targets. It surfaces model posture and target-level metrics in the UI and governance data. That is an important product behavior: weaker model areas should be visible, not disguised behind a single aggregate score.

RVP **Radiology Value Pipeline**
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

Human review remains required
Assistive output only
Chest X-ray + Brain MRI

Clinical operations workspace

Chest X-Ray 2 Views • processed • CR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**ACC-661D5B5B**	**baseline**	**3:04:20 PM**

Imaging review

Preview image, modality context, and viewer launch controls.

Launch OHIF | Launch Orthanc

Study preview | Preview only | Source

Integration status
OHIF sidecar: offline
Orthanc: offline
Launch mode: external_link
Preview artifact: available

Viewer sidecars are not running on this machine, so the review surface stays on the local diagnostic preview instead of embedding a broken frame.

Review metrics

Modality	Chest X-ray
Mode	baseline
Runs	8
Findings	1
Feedback	1
Discrepancies	1
Agent runs	4
Preview	available

AI synthesis

Urgent

Findings, modality-aware diagnostics, and workflow trace.

Candidate findings

focal airspace opacity pattern
mild • confidence 0.65 • abnormality_detection_service

Model posture

Model bundle	Validation AUC	Held-out test AUC	Ensemble members	Positive targets
compact-cnn-v4-ensemble-2	**0.760**	**0.747**	**2**	**1**
compact_cnn_ensemble	ChestMNIST	Deployment benchmark	Current inference bundle	Current study output

Target	Probability	Threshold / Mode	Prediction	Val AUC	Test AUC	Demo truth
cardiomegaly	0.008	0.725	negative	0.872	0.860	negative compact-cnn-v4-ensemble-2
infiltration	0.654	0.650	positive	0.655	0.666	negative infiltration-specialist-v2-ensemble-2

Urgency rationale

Escalated due to focal airspace finding.

Draft findings

right focal airspace opacity pattern (confidence 0.65)

Draft impression

AI candidate findings are listed for radiologist review. Suggested urgency tier: urgent. No prior comparison available in local demo store.

Specialist service ledger

Service	Status	Latency	Mode
representation_service	Success	5 ms	Baseline
view_quality_service	Success	5 ms	Baseline
abnormality_detection_service	Success	5 ms	Baseline
urgency_scoring_service	Success	5 ms	Baseline
prior_context_service	Success	5 ms	Baseline
explanation_service	Success	5 ms	Baseline
report_draft_service	Success	5 ms	Baseline
discrepancy_service	Success	5 ms	Baseline

Workflow agent ledger

Agent	Status	Latency	Prompt
explanation_review_agent	Success	5 ms	8b61c4f2-9914-4182-b7d4-c256c5b886af
report_assist_agent	Success	5 ms	ae77fd4c-084b-4014-a54a-74d2c25c87cc
discrepancy_review_agent	Success	5 ms	b6e79d43-4409-41c0-b7b8-286f53cd6deb
discrepancy_review_agent	Success	5 ms	b6e79d43-4409-41c0-b7b8-286f53cd6deb

Artifacts

source_preview_png
/artifacts/studies/1d2c9c32-391e-4113-b902-53a1a3824730/source-preview.png

Reporting workspace

Human review, edits, and final report persistence.

Mark Helpful | Mark Unhelpful

Final report

Chest X-ray benchmark labels do not indicate a high-priority target finding. Radiologist review completed with human-authored final impression. AI urgency suggestion was urgent. AI draft was treated as

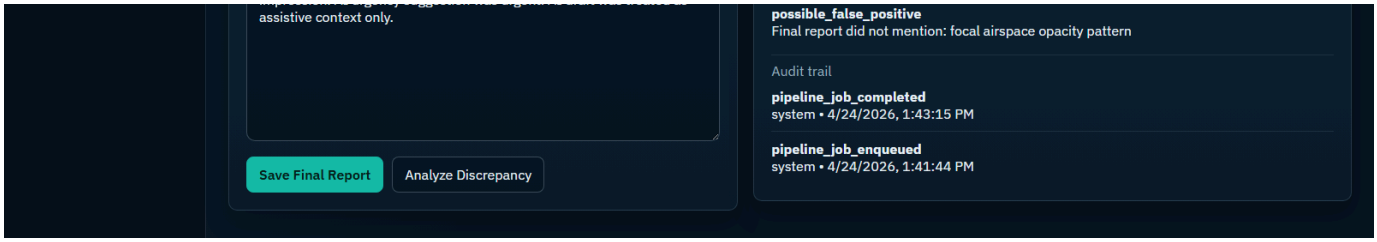
Review event stream

Feedback, discrepancy notes, and audit trail.

Feedback

accept_finding
demo.radiologist@example.local • Bulk demo simulated radiologist accepted the candidate finding.

Discrepancies



12.2 Brain MRI Advanced Imaging Path

Brain MRI context and segmentation bundle

Runtime bundle: brain-mri-context-seg-v2

Dataset: public Figshare Brain Tumor MRI benchmark

Validation macro AUC: 0.9940

Held-out test macro AUC: 0.9828

Segmentation validation Dice: 0.3621

Segmentation held-out test Dice: 0.3556

Brain MRI is the “advanced imaging AI” signal in the project. It shows that the platform can support more than 2D X-ray classification.

The current MRI path includes:

- Multi-class tumor phenotype classification.
- Context-aware slice classification.
- Segmentation-backed localization.
- Preview, mask, reference overlay, and saliency artifacts.
- Patient-level volumetric 3D corroboration.
- Modality-aware findings, urgency rationale, report draft text, and dashboard reporting.

Current brain MRI target results:

Target	Test AUC
Meningioma	0.9787
Glioma	0.9831
Pituitary	0.9865

Target	Test AUC
Macro average	0.9828

RVP

Radiology Value Pipeline
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

Human review remains required

Assistive output only

Chest X-ray + Brain MRI

Clinical operations workspace

Brain MRI T1 Post-Contrast • processed • MR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**MR-1B88A9E1**	**baseline**	**3:01:57 PM**

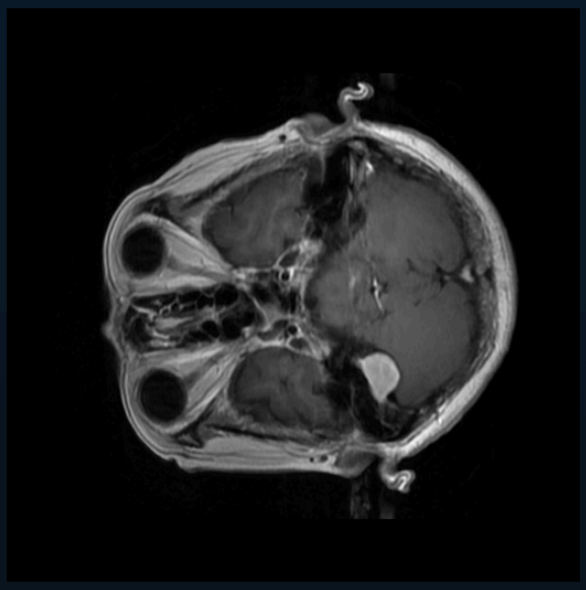
Imaging review

Launch OHIF | Launch Orthanc

Preview image, modality context, and viewer launch controls.

Study preview | Preview only

Source | Reference overlay | Reference mask | AI saliency | AI segmentation | AI mask | Context strip | 3D projection



Integration status

OHIF sidecar: offline

Orthanc: offline

Launch mode: external_link

Preview artifact: available

Viewer sidecars are not running on this machine, so the review surface stays on the local diagnostic preview instead of embedding a broken frame.

Review metrics

Modality	Brain MRI
Mode	baseline
Runs	8
Findings	4
Feedback	1
Discrepancies	1
Agent runs	4
Preview	available

AI synthesis

Routine

Findings, modality-aware diagnostics, and workflow trace.

Candidate findings

intracranial enhancing mass lesion candidate
high • confidence 1.00 • abnormality_detection_service

dural-based extra-axial mass phenotype
moderate • confidence 1.00 • abnormality_detection_service

patient-level 3D MRI phenotype classification available
moderate • confidence 1.00 • abnormality_detection_service

segmentation-backed lesion localization available
high • confidence 0.50 • abnormality_detection_service

Model posture

Model bundle
brain-mri-context-seg-v2
context_ensemble_plus_slice_segmentation_with_3d_volume_support

Validation AUC
0.994
Figshare Brain Tumor MRI

Held-out test AUC
0.983
Deployment benchmark

Predicted phenotype
meningioma
context-window-3 primary • confidence 0.999

3D volume AUC
0.981
5 slices

Segmentation Dice
0.738
Current demo overlap

Lesion extent
0.6%
Volume burden 0.5%

Context window
3
Slices 56 • 56 • 57

Target	Probability	Threshold / Mode	Prediction	Val AUC	Test AUC	Demo truth
glioma	0.001	context-window-3	negative	0.993	0.983	negative context-window-3
meningioma	0.999	context-window-3	positive	0.989	0.969	positive context-window-3
pituitary_tumor	0.000	context-window-3	negative	1.000	0.991	negative context-window-3

Urgency rationale

MRI phenotype is suspicious for an intracranial mass lesion, but the current patient-volume confidence and localization burden support routine subspecialty review rather than acute escalation.

Draft findings

right intracranial enhancing mass lesion candidate (confidence 1.00); right dural-based extra-axial mass phenotype (confidence 1.00); right patient-level 3D MRI phenotype classification available (confidence 1.00); right segmentation-backed lesion localization available (confidence 0.50)

Draft impression

AI candidate MRI findings are listed for neuroradiology review. Suggested urgency tier: routine. No prior volumetric comparison is available in the local MRI demo store.

Specialist service ledger

Service	Status	Latency	Mode
abnormality_detection_service	Success	5 ms	Baseline
discrepancy_service	Success	5 ms	Baseline
explanation_service	Success	5 ms	Baseline
prior_context_service	Success	5 ms	Baseline
report_draft_service	Success	5 ms	Baseline
representation_service	Success	5 ms	Baseline
urgency_scoring_service	Success	5 ms	Baseline
view_quality_service	Success	5 ms	Baseline

Workflow agent ledger

Agent	Status	Latency	Prompt
explanation_review_agent	Success	5 ms	8b61c4f2-9914-4182-b7d4-c256c5b886af
report_assist_agent	Success	5 ms	ae77fd4c-084b-4014-a54a-74d2c25c87cc

discrepancy_review_agent	Success	5 ms	b6e79d43-4409-41c0-b7b8-286f53cd6deb
discrepancy_review_agent	Success	5 ms	b6e79d43-4409-41c0-b7b8-286f53cd6deb

Artifacts

source_preview_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/source-preview.png

reference_mask_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/reference-mask.png

reference_overlay_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/reference-overlay.png

ai_saliency_overlay_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/ai-saliency-overlay.png

ai_segmentation_mask_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/ai-segmentation-mask.png

ai_segmentation_overlay_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/ai-segmentation-overlay.png

ai_context_strip_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/ai-context-strip.png

ai_volume_projection_png
/artifacts/studies/744364ec-c32e-4c53-8e22-4cd7713e0b63/ai-volume-projection.png

Reporting workspace

Human review, edits, and final report persistence.

Mark Helpful | Mark Unhelpful

Final report

Brain MRI public benchmark phenotype: brain MRI phenotype. Radiologist review completed with human-authored final impression. AI urgency suggestion was routine. AI draft was treated as assistive context only.

Save Final Report | Analyze Discrepancy

Review event stream

Feedback, discrepancy notes, and audit trail.

Feedback

accept_finding
demo.radiologist@example.local • Bulk demo simulated radiologist accepted the candidate finding.

Discrepancies

possible_false_positive
Final report did not mention: intracranial enhancing mass lesion candidate, dural-based extra-axial mass phenotype, patient-level 3D MRI phenotype classification available, segmentation-backed lesion localization available

Audit trail

pipeline_job_completed
system • 4/24/2026, 1:44:27 PM

pipeline_job_enqueued
system • 4/24/2026, 1:41:44 PM

12.3 Patient-Level 3D MRI Corroboration

Brain MRI patient-level 3D corroboration

Runtime bundle: brain-mri-volume3d-v1

Volume shape: 32 x 96 x 96

Hybrid validation macro AUC: 0.9941

Hybrid held-out test macro AUC: 0.9808

Validation accuracy: 0.9714

Held-out test accuracy: 0.8571

The 3D path is valuable because it moves the MRI workflow beyond isolated 2D slices. A real neuroradiology workflow reasons across volumes and context. The current path stacks patient-level volumes from the public benchmark, trains a compact 3D model, and blends it with the stronger 2D context model for a hybrid corroboration score.

This is still not a multi-sequence BraTS-style workflow. It is, however, an important architectural proof that the same product can support:

- 2D high-throughput radiography.
- 2D context-aware MRI classification.
- Segmentation-backed localization.
- 3D patient-level corroboration.
- Shared governance, feedback, and dashboarding across modalities.

13 Review Workspace

The browser review workspace is designed to feel like a clinical operations surface rather than a model demo page.

The study review page shows:

- Modality and accession context.
- Imaging preview and sidecar viewer launch controls.
- Clear OHIF and Orthanc status.
- Model bundle, validation AUC, test AUC, ensemble size, and target-level probabilities.
- Candidate findings with body region, severity, confidence, source service, status, and evidence references.

- Urgency rationale.
- Draft findings and impression text.
- Specialist service ledger.
- Workflow agent ledger.
- Artifact links.
- Reporting workspace for final report text.
- Feedback buttons.
- Discrepancy events.
- Audit trail.

The UI is intentionally explicit. If a sidecar service is offline, it says so. If an artifact is local, it shows the local fallback. If the AI is assistive only, the copy says so.

14 Operations, Quality, and Governance Dashboards

Power BI is included because the product is not only for model builders. A radiology AI platform needs visibility for operations leaders, clinical quality teams, and governance committees.

RVP

Radiology Value Pipeline
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

- Human review remains required
- Assistive output only
- Chest X-ray + Brain MRI

Clinical operations workspace

Brain MRI T1 Post-Contrast • processed • MR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**MR-1B88A9E1**	**baseline**	**2:55:52 PM**

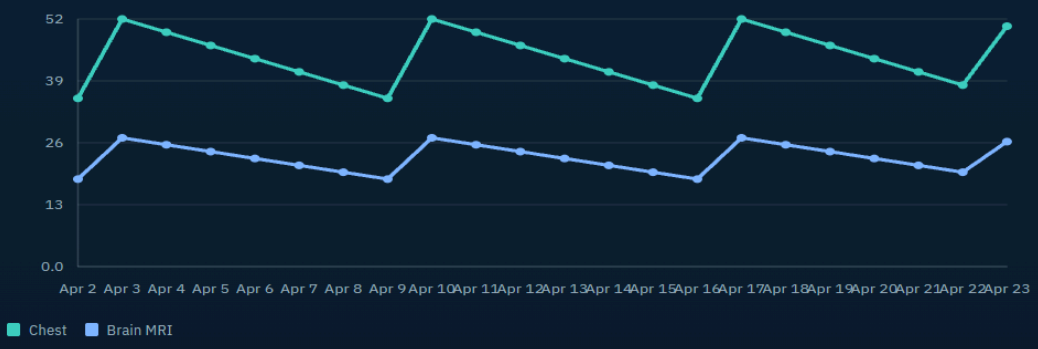
Operations dashboard

Queue posture, throughput, and latency watch.

Queue backlog	Processed today	Chest studies	MRI studies
0	**76**	**50**	**26**
Apr 23	Apr 23	Apr 23	Apr 23

Volume trend

Processed and urgent study flow across the demo history window.



Operational metric board

Latest snapshot values, normalized for quick scanning.

Queue backlog	0
Urgent cases	30
Failed runs	0
Registered studies	76
Latency p50	5 ms
Latency p90	5 ms

RVP

Radiology Value Pipeline
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

- Human review remains required
- Assistive output only
- Chest X-ray + Brain MRI

Clinical operations workspace

Brain MRI T1 Post-Contrast • processed • MR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**MR-1B88A9E1**	**baseline**	**2:55:52 PM**

Clinical quality dashboard

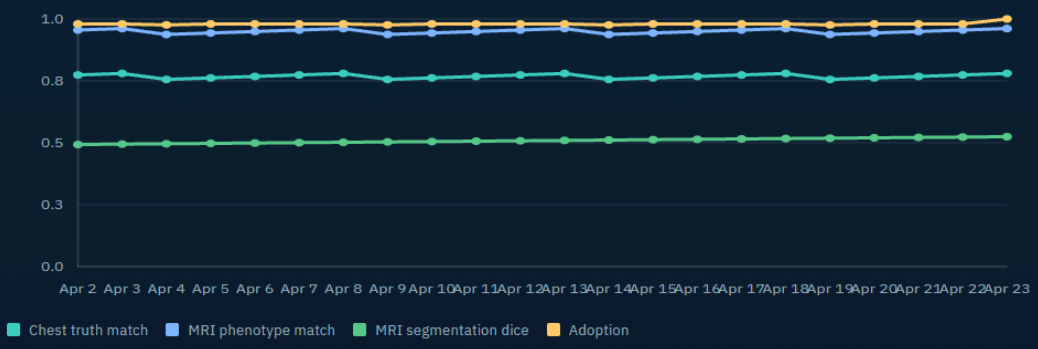
Concordance, adoption, and discrepancy behavior.

Chest truth match	MRI phenotype match
78.0%	**96.2%**
Apr 23	Apr 23

MRI segmentation dice	Draft adoption
0.525	**100.0%**
Apr 23	Apr 23

Quality trend

Rate metrics plotted across the seeded demo timeline.



Quality metric board

Latest quality metrics with shared scale treatment.

Candidate findings	129
Feedback events	48
Discrepancy events	76
Chest positive precision	28.0%
Chest positive recall	63.6%
Abstain rate	0.0%

RVP **Radiology Value Pipeline**
Radiology workflow platform

WORKSPACE

- Worklist
- Study Review
- Operations
- Clinical Quality
- Governance
- Power BI

GUARDRAILS

Human review remains required
Assistive output only
Chest X-ray + Brain MRI

Clinical operations workspace

Brain MRI T1 Post-Contrast • processed • MR

Pipeline mode: baseline | Refresh | Run Pipeline | API Docs

Environment	Active study	Pipeline mode	Last refresh
Live local stack	**MR-1B88A9E1**	**baseline**	**2:55:52 PM**

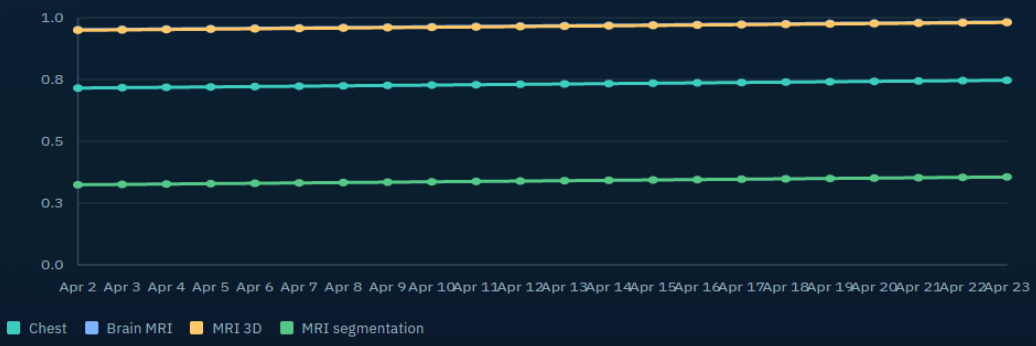
Governance dashboard

Versioned models, workflow agents, and deployment posture.

Chest test AUC	Brain MRI test AUC	MRI 3D test AUC	MRI segmentation dice	MRI glioma AUC
0.747	**0.983**	**0.981**	**0.356**	**0.983**
Apr 23	Apr 23	Apr 23	Apr 23	Apr 23

Model performance trend

Baseline and specialist AUC history for the demo environment.



Governance metric board

Active governance, MRI readiness, and deployment watch metrics.

Agent runs	304
Signed reports	76
MRI validation AUC	0.994
MRI 3D val AUC	0.994
MRI val dice	0.362
Chest specialist gain	+0.032
Unhelpful feedback	0

Model inventory

Registry alignment and deployment mode.

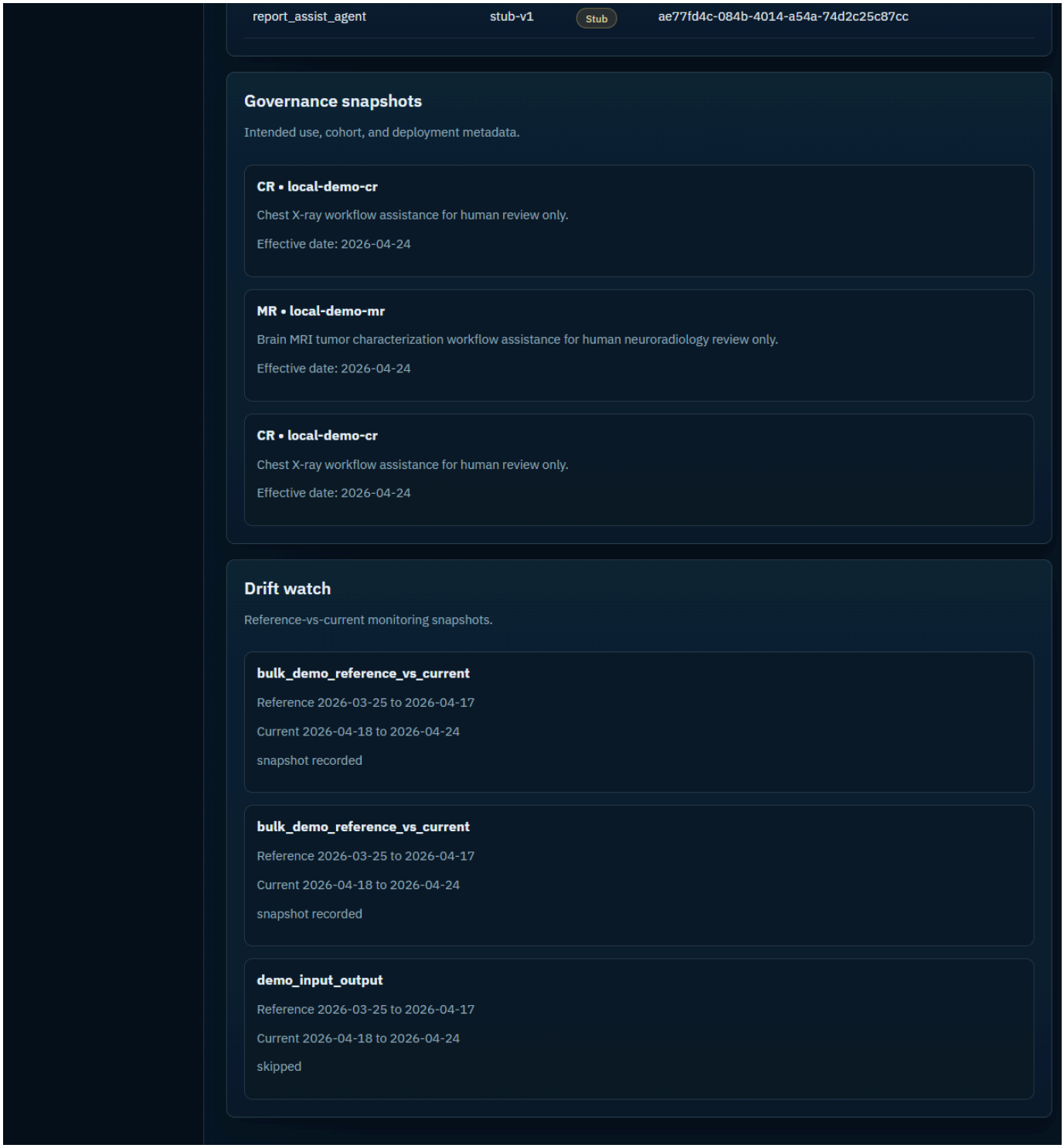
Service	Modality	Version	Mode	Intended use
abnormality_detection_service	Study	brain-mri-context-seg-v2	Baseline	Not set
abnormality_detection_service	Study	compact-cnn-v4-	Baseline	Not set

		ensemble-2		
abnormality_detection_service	Study	stub-cr-v1	Stub	Not set
abnormality_detection_service	Study	stub-mr-v1	Stub	Not set
discrepancy_service	Study	baseline-mr-v1	Baseline	Not set
discrepancy_service	Study	baseline-cr-v1	Baseline	Not set
discrepancy_service	Study	stub-mr-v1	Stub	Not set
discrepancy_service	Study	stub-cr-v1	Stub	Not set
explanation_service	Study	baseline-mr-v1	Baseline	Not set
explanation_service	Study	baseline-cr-v1	Baseline	Not set
explanation_service	Study	stub-mr-v1	Stub	Not set
explanation_service	Study	stub-cr-v1	Stub	Not set
prior_context_service	Study	baseline-mr-v1	Baseline	Not set
prior_context_service	Study	baseline-cr-v1	Baseline	Not set
prior_context_service	Study	stub-cr-v1	Stub	Not set
prior_context_service	Study	stub-mr-v1	Stub	Not set
report_draft_service	Study	baseline-mr-v1	Baseline	Brain MRI tumor characterization workflow assistance for human neuroradiology review only.
report_draft_service	Study	baseline-cr-v1	Baseline	Chest X-ray workflow assistance for human review only.
report_draft_service	Study	stub-cr-v1	Stub	Not set
report_draft_service	Study	stub-mr-v1	Stub	Not set
representation_service	Study	baseline-mr-v1	Baseline	Not set
representation_service	Study	baseline-cr-v1	Baseline	Not set
representation_service	Study	stub-cr-v1	Stub	Not set
representation_service	Study	stub-mr-v1	Stub	Not set
urgency_scoring_service	Study	baseline-mr-v1	Baseline	Not set
urgency_scoring_service	Study	baseline-cr-v1	Baseline	Not set
urgency_scoring_service	Study	stub-cr-v1	Stub	Not set
urgency_scoring_service	Study	stub-mr-v1	Stub	Not set
view_quality_service	Study	baseline-mr-v1	Baseline	Not set
view_quality_service	Study	baseline-cr-v1	Baseline	Not set
view_quality_service	Study	stub-mr-v1	Stub	Not set
view_quality_service	Study	stub-cr-v1	Stub	Not set

Agent inventory

Workflow agents and prompt linkage.

Agent	Version	Mode	Prompt
discrepancy_review_agent	stub-v1	Stub	b6e79d43-4409-41c0-b7b8-286f53cd6deb
explanation_review_agent	stub-v1	Stub	8b61c4f2-9914-4182-b7d4-c256c5b886af

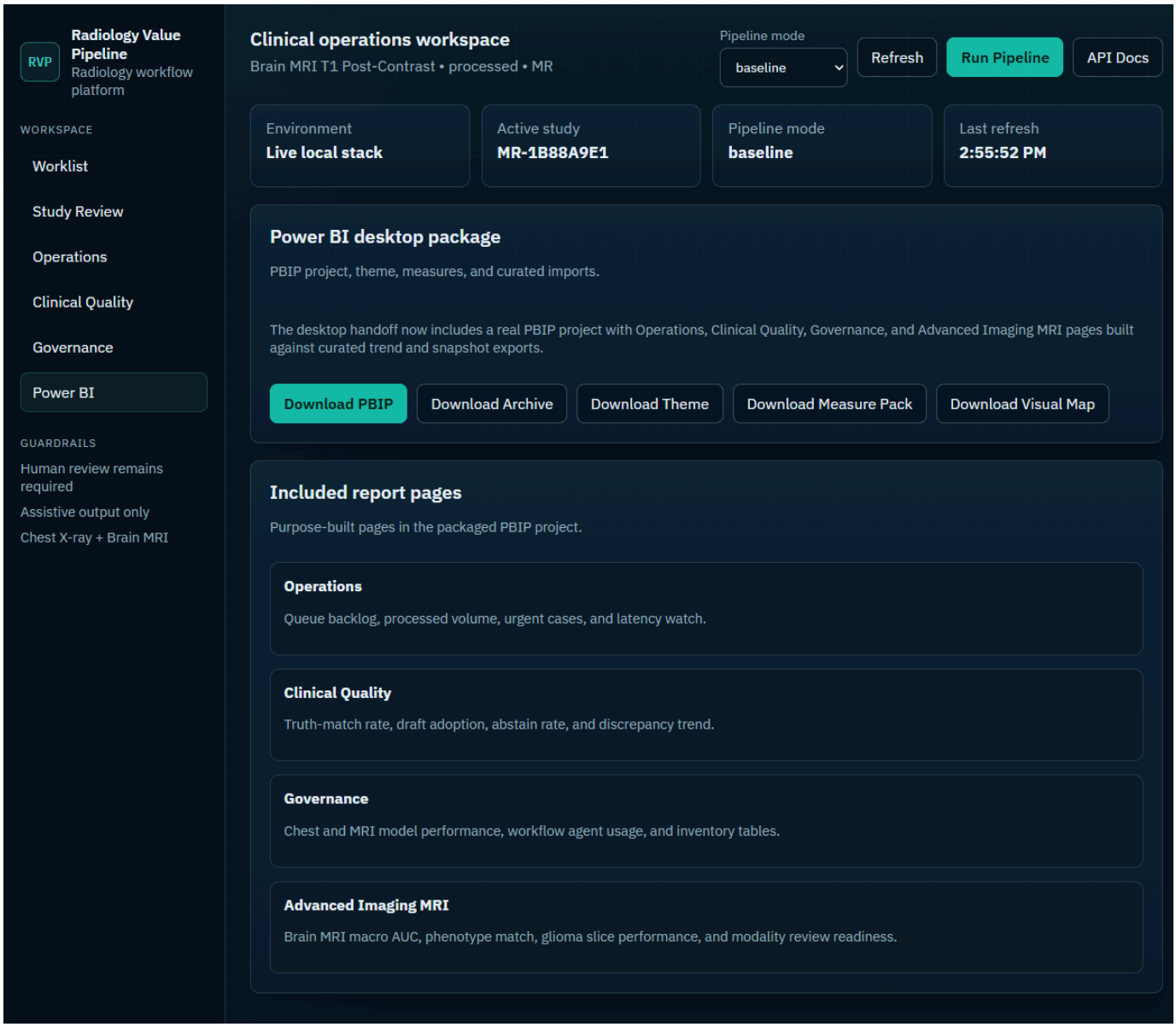


14.1 Power BI Report Package

The Power BI package is designed around curated exports from dashboard marts. The report includes:

- Executive overview.
- Operational volume and latency.
- Clinical quality and model performance.
- Governance inventory and drift posture.
- Model and agent lineage.

The project includes a PBIP handoff package, CSV exports, and a browser download surface.



14.2 Embedded Power BI Export

The exported Power BI PDF is embedded below for HTML renderers that support PDF objects. The individual pages are also extracted as PNG images for portability.

Open the exported Power BI PDF at assets/powerbi/RadiologyValuePipelineDashboard.pdf.

Queue backlog

0.00

Studies processed

76.00

Chest studies processed

50.00

Brain MRI studies processed

26.00

Chest study volume trend

Brain MRI volume trend

Operational trend table

event_date	queue_counts	cr_studies_processed	mr_studies_processed	urgent_cases	failed_runs	latency_p90_ms
Friday, April 03, 2026	1.00	35.00	18.20	21.00	1.00	5.52
Saturday, April 04, 2026	0.00	51.50	26.78	30.90	0.00	5.40
Sunday, April 05, 2026	3.00	48.75	25.35	29.25	3.00	5.78
Monday, April 06, 2026	2.00	46.00	23.92	27.60	2.00	5.65
Tuesday, April 07, 2026	1.00	43.25	22.49	25.95	1.00	5.52

Chest truth match rate

0.78

Brain MRI phenotype match

0.96

Draft assist adoption

1.00

Brain MRI high confidence

0.96

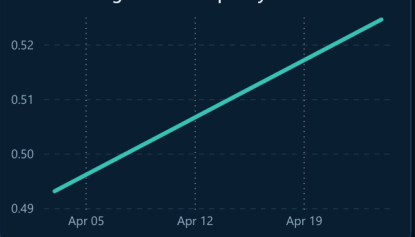
Brain MRI ref dice

0.52

Chest truth match trend

Brain MRI phenotype match trend

Brain MRI segmentation quality trend



Clinical quality trend table

event_date	candidate_findings	feedback_events	discrepancy_events	draft_assist_adoption_rate	baseline_truth_match_rate	brain_mri_phenotype_match_rate	brain_mri_high
Friday, April 03, 2026	90.30	33.60	53.20	0.98	0.77	0.96	
Saturday, April 04, 2026	132.87	49.44	78.28	0.98	0.78	0.96	
Sunday, April 05, 2026	125.78	46.80	74.10	0.98	0.76	0.94	
Monday, April 06, 2026	118.68	44.16	69.92	0.98	0.76	0.94	
Tuesday, April 07, 2026	111.59	41.52	65.74	0.98	0.77	0.95	

15 What Makes This Portfolio Project Different

Many AI portfolio projects stop at a notebook. This one builds the product shell around the model.

The work demonstrates:

- Platform architecture, not just experimentation.
- Real database modeling with lineage and governance entities.
- Worker orchestration without distributed complexity.
- Multi-modality support across chest X-ray and brain MRI.
- Real benchmark-trained models where the project needs credibility.
- Browser review workflow with human feedback and discrepancy handling.
- Power BI reporting for operations, quality, and governance.
- Auditability and versioning as first-class features.
- A clear boundary between real model components and scaffolded future integrations.

This is the difference between “I trained a model” and “I understand what it takes to make imaging AI useful inside a clinical workflow.”

16 Clinical Workflow Design

The platform follows a human-in-the-loop path:

1. A study is registered from a DICOM gateway or demo adapter.
2. The study metadata, series, instances, order context, and artifact references are persisted.
3. A pipeline job is enqueued.
4. The worker runs ordered specialist services.
5. Structured findings and urgency suggestions are persisted.
6. A draft report is rendered from structured findings.
7. A clinician reviews the study and AI suggestions.
8. Feedback is recorded as append-only events.
9. A final report is saved as clinician-authored or clinician-edited text.
10. Discrepancy analysis compares AI outputs with the final report.
11. Dashboard marts expose adoption, latency, override, discrepancy, drift, and governance signals.

This workflow avoids the most dangerous pattern in medical AI demos: producing a confident paragraph with no lineage, no feedback, no uncertainty, and no accountability.

17 Report Assist Design

The report assist layer does not merely output a free-form paragraph. It produces structured data first and text second.

Canonical finding shape:

```
{
  "body_region": "right lung",
  "finding_name": "focal airspace opacity pattern",
  "laterality": "right",
  "severity": "mild",
  "confidence": 0.65,
  "evidence_ref": "source_preview_png",
  "source_service": "abnormality_detection_service",
  "status": "candidate",
  "uncertainty_flag": false,
  "abstain_reason": null
}
```


Draft text is then generated from the structured payload:

```
Draft findings:
right focal airspace opacity pattern (confidence 0.65)

Draft impression:
AI candidate findings are listed for radiologist review.
Suggested urgency tier: urgent.
No prior comparison available in local demo store.
```

This structure is important because downstream systems can reason about a finding object, but they cannot safely reason about a free-form paragraph alone.

18 Feedback Loop

Feedback is append-only and typed. The system supports:

- Accept finding.
- Reject finding.
- Edit finding.
- Override urgency.
- Mark helpful or unhelpful.
- Submit discrepancy note.
- Submit issue type such as false positive, false negative, bad localization, or hallucinated text.

This creates a learning system without pretending to retrain automatically. Feedback is stored as evidence for review, analytics, and future model improvement.

19 Governance and Monitoring

The governance layer is one of the strongest parts of the project because it treats model deployment as an accountable lifecycle.

Governance snapshots store:

- Intended use.
- Modality.
- Supported populations and exclusions.
- Training data provenance.
- Validation summary.
- Known limitations.
- Rollback plan.
- Effective date.
- Deployment cohort.
- Threshold configuration.

Monitoring stores:

- Model lineage.
- Prompt lineage.
- Agent lineage.
- Latency.
- Drift summaries.
- Override rate.
- Abstain rate.
- Subgroup quality metrics.
- Unresolved incidents.
- Audit completeness.

This is directly aligned with the direction of modern clinical AI governance: transparency, versioning, real-world monitoring, and clear human accountability.

20 API Surface

The API is intentionally simple and reviewable.

Endpoint	Purpose
`GET /healthz`	Health check
`POST /studies/register`	Register study metadata

Endpoint	Purpose
`GET /studies/{study_id}`	Retrieve study detail
`POST /studies/{study_id}/run-pipeline`	Enqueue or run model pipeline
`GET /studies/{study_id}/findings`	Retrieve structured findings
`GET /studies/{study_id}/report-draft`	Retrieve draft report payload
`POST /studies/{study_id}/feedback`	Record feedback
`POST /studies/{study_id}/final-report`	Save final report
`POST /studies/{study_id}/discrepancy/analyze`	Analyze discrepancy
`GET /studies/{study_id}/review-context`	Retrieve full UI review context
`GET /dashboards/ops`	Operational dashboard summary
`GET /dashboards/quality`	Clinical quality dashboard summary
`GET /dashboards/governance`	Governance dashboard summary
`GET /governance/models`	Model inventory
`GET /governance/agents`	Agent inventory
`GET /governance/drift`	Drift snapshot summary

Example review context response shape:


```
{
  "study": {
    "id": "1d2c9c32-391e-4113-b902-53a1a3824730",
    "modality": "CR",
    "description": "Chest X-Ray 2 Views",
    "status": "processed"
  },
  "model_posture": {
    "model_bundle": "compact-cnn-v4-ensemble-2",
    "validation_auc": 0.7597,
    "held_out_test_auc": 0.7470
  },
  "findings": [],
  "urgency": {},
  "draft_report": {},
  "agent_runs": [],
```

```
  "audit_events": []
}
```

21 Security and Privacy Posture

The MVP is local-first and privacy-safe by default:

- No PHI is required for the demo path.
- Synthetic and public benchmark data are used for demonstration.
- DICOM binaries are not stored in Postgres.
- Study artifacts are stored in MinIO or filesystem abstraction.
- Site scoping is included for future RLS.
- Local auth headers are scaffolded but enterprise identity is deferred.
- Audit events record key workflow actions.
- The system does not make clinical deployment claims.

22 What Is Real vs Stubbed

The project is honest about maturity.

Area	Current state
Study registration	Real local workflow and API
PostgreSQL schema	Real migrations and models
Worker orchestration	Real DB-backed queue pattern
Chest X-ray model	Real ChestMNIST compact CNN ensemble baseline
Infiltration specialist	Real target-specific baseline improvement
Brain MRI model	Real public Figshare benchmark model path
Brain MRI 3D corroboration	Real patient-level stacked-volume model path
Findings persistence	Real structured persistence
Draft report rendering	Real deterministic renderer, not a clinical LLM
Feedback events	Real append-only persistence
Discrepancy analysis	Simplified deterministic comparison

Area	Current state
Governance snapshots	Real tables and surfaces, benchmark-scoped values
Drift snapshots	Scaffolded and demo-populated
OHIF integration	Sidecar and launch contract, deep embedding deferred
Orthanc integration	DICOMweb adapter contract, live sidecar depends on Docker
Enterprise identity	Deferred
Clinical validation	Not performed

23 What I Want A Clinical Engineering Team To Notice

I built this project to show how I think about clinical software, not just how I train a model.

The model matters, but the system around the model matters just as much. In a real radiology setting, a good prediction is only useful if the team can see where it came from, understand what version produced it, review the evidence, correct it, audit it later, and measure whether it helped or hurt the workflow.

That is why I put so much weight on the parts that are often missing from AI demos:

- The radiologist remains the author of record.
- AI outputs are stored as structured candidates, not as untraceable narrative.
- Every run links back to a model version, prompt version, input payload, output payload, and audit record.
- Feedback and discrepancies are preserved instead of being treated as one-off UI actions.
- Dashboards are built from SQL marts so operational, quality, and governance questions can be answered outside the application.
- The architecture stays simple enough for another engineer to inherit.

What I want a reviewer to see is that I care about the whole product surface: the imaging edge, the database, the model contracts, the review workflow, the reporting layer, the failure modes, and the documentation. I did not want this to be a notebook that looks impressive for five minutes. I wanted it to feel like the beginning of a system a serious team could critique, extend, and eventually harden.

24 How I Would Walk A Reviewer Through It

If we were walking through the repository together, I would start with the workflow instead of the model. I would show how a study enters the system, how the pipeline job is created, how each specialist service writes a run record, and how the review page makes the model output inspectable.

Then I would show the database. That is where the project becomes more than a demo. The schema captures studies, model versions, prompt versions, model runs, findings, urgency scores, report drafts, final reports, feedback, discrepancies, audit events, governance snapshots, drift snapshots, and dashboard facts. It is designed so a clinical team can ask: what happened, who reviewed it, what changed, and how often is this pattern occurring?

After that I would show the two imaging paths. Chest X-ray demonstrates the high-throughput workflow and the practical challenge of weaker targets like infiltration. Brain MRI demonstrates the more advanced imaging path: multi-class tumor phenotype classification, segmentation-backed localization, and a patient-level 3D corroboration model. I would be explicit that these are benchmark-trained models and not clinically validated devices.

Finally, I would show the dashboards and governance surfaces. My goal there was to make the product legible to more than one audience. Engineers can inspect runs and versions. Clinicians can review findings and feedback. Operations leaders can see throughput and latency. Governance reviewers can see model inventory, prompts, drift records, limitations, and audit completeness.

25 Next Milestones

The next serious increments would be:

1. Harden DICOM ingestion against a live Orthanc instance and test OHIF launch URLs with real DICOMweb studies.
2. Replace selected deterministic services with stronger MONAI Bundle-backed implementations.
3. Add a richer BraTS-style multi-sequence MRI path with volumetric segmentation and clinically meaningful MRI report structure.
4. Add calibrated uncertainty, abstention thresholds, and subgroup-level evaluation.
5. Add stronger NLP discrepancy analysis with explicit prompt governance and human review.
6. Expand identity from local headers to enterprise SSO.
7. Add scheduled drift jobs and alert review workflows.
8. Add a read replica or export pipeline for larger Power BI deployments.
9. Run a silent-mode evaluation against de-identified retrospective cases under appropriate governance.

26 Local Run Path

The repo includes a one-command local path:

```
powershell -ExecutionPolicy Bypass -File scripts/start_local_mvp.ps1
```

For the already-seeded local environment, the fast path is:

```
powershell -ExecutionPolicy Bypass -File scripts/start_local_mvp.ps1 -SkipSeed -SkipDemo
```

Then open:

```
http://127.0.0.1:8000/
```

API docs:

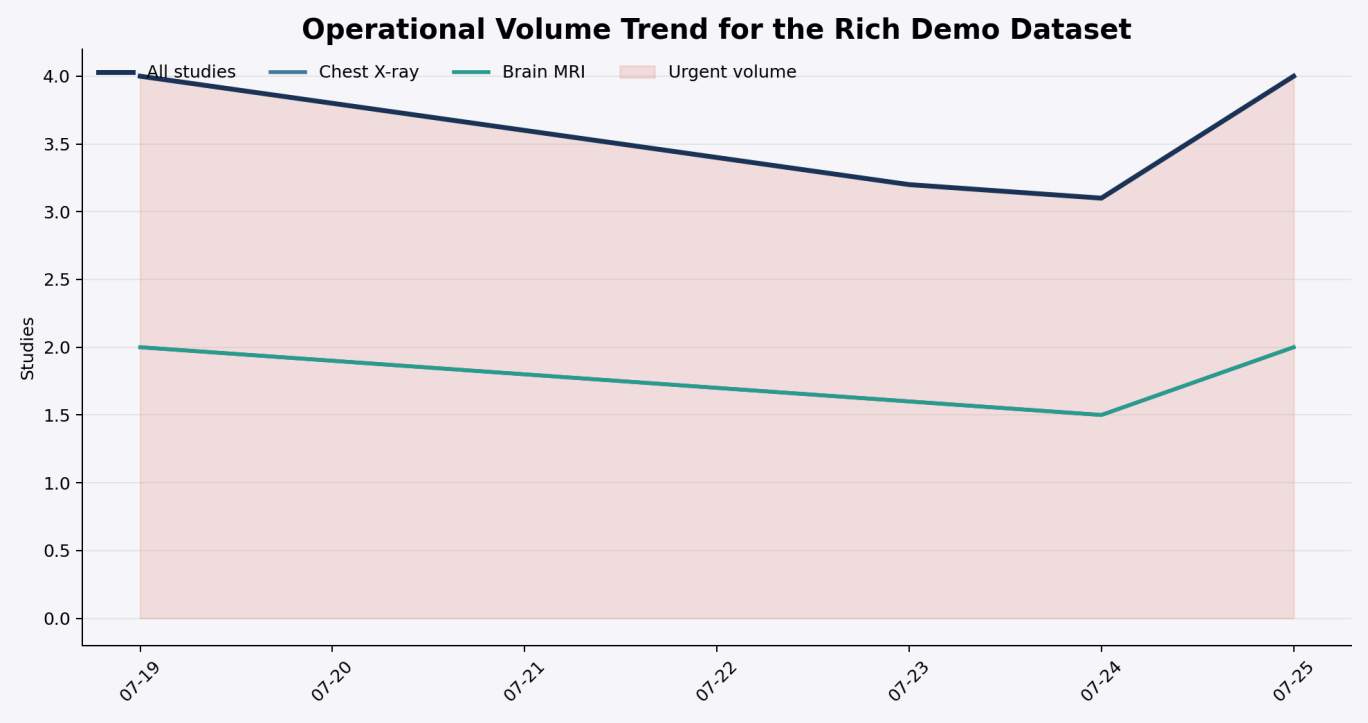
```
http://127.0.0.1:8000/docs
```

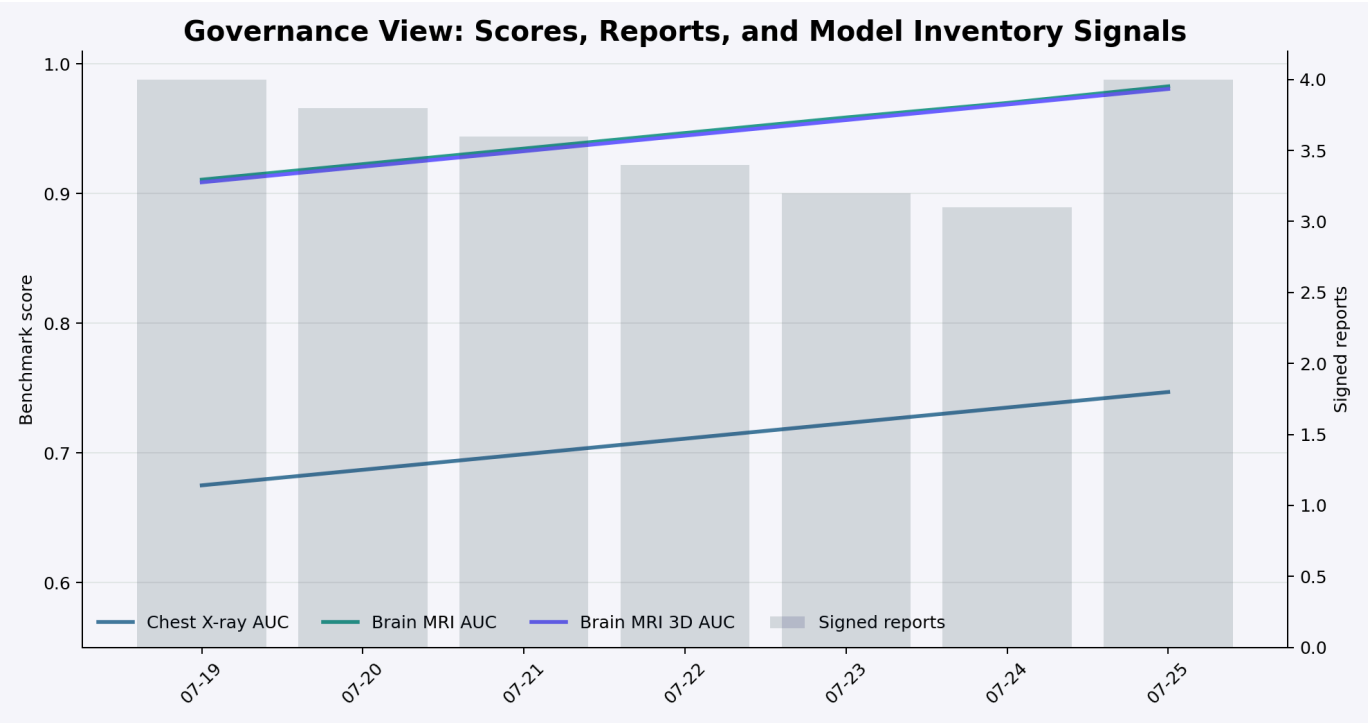
27 Appendix: Dashboard Data Export Shape

The Power BI package consumes curated exports such as:

```
dashboards/powerbi/exports/ops_snapshot.csv
dashboards/powerbi/exports/quality_snapshot.csv
dashboards/powerbi/exports/governance_snapshot_latest.csv
dashboards/powerbi/exports/ops_trend.csv
dashboards/powerbi/exports/quality_trend.csv
dashboards/powerbi/exports/governance_trend.csv
dashboards/powerbi/exports/model_versions.csv
dashboards/powerbi/exports/agent_versions.csv
dashboards/powerbi/exports/agent_runs.csv
```

This makes the handoff portable. A reviewer can understand the dashboard model even without connecting to the local PostgreSQL database.





28 Appendix: SQL Mart Excerpts

Operational dashboard:

```
select
  event_date,
  site_code,
  queue_counts,
  studies_registered,
```

```
  studies_processed,
  urgent_cases,
  failed_runs,
  latency_p50_ms,
  latency_p90_ms
from dashboard_ops_daily;
```

Clinical quality dashboard:

```
select
  event_date,
  site_code,
  candidate_findings,
  feedback_events,
  helpful_feedback_rate,
  discrepancy_events,
  draft_assist_adoption_rate,
  abstain_rate,
  baseline_truth_match_rate,
  baseline_positive_recall_rate,
  baseline_positive_precision_rate
from dashboard_quality_daily;
```

Governance dashboard:

```
select
  event_date,
  site_code,
  governance_snapshots,
  baseline_runs,
  agent_runs,
  signed_reports,
  unhelpful_feedback,
  baseline_validation_auc,
  baseline_test_auc,
  infiltration_specialist_test_auc,
  infiltration_specialist_gain_auc
from dashboard_governance_daily;
```

29 Appendix: Product Boundaries

This project is meant to be impressive because it is disciplined.

It does not claim:

- FDA clearance.
- Clinical deployment readiness.
- Replacement of radiologists.
- Autonomous diagnosis.
- Real hospital integration.
- Production security hardening.
- Prospective clinical validation.

It does claim:

- A serious architecture for radiology workflow AI.
- Real local execution.
- Real schema design.
- Real model lineage and governance surfaces.
- Real benchmark-trained model paths.
- Real browser review workflow.
- Real Power BI reporting handoff.
- A credible path from portfolio MVP to clinical engineering platform.

30 Closing

I built Radiology Value Pipeline because radiology AI should be measured by more than a model score. It should be measured by whether it can fit into real clinical work, preserve accountability, learn from feedback, and make diagnostic systems more reliable over time.

That is the project I wanted to show: a platform that respects the clinician, respects the patient, and treats AI as one part of a larger care-delivery system.