HPH Framework: Research Project Plan

Ivan Ogasawara[1]

[1]Independent Researcher, IGDORE

November 2025

1. Background and Rationale

High-quality clinical care hinges on continuity and accessibility of information across the entire encounter: screening, diagnosis, treatment, prescription, and follow-up. Despite major gains in electronic health record (EHR) adoption over the past decade, fragmentation and workflow misalignment persist in practice. For example, U.S. office-based physician EHR adoption reached 88.2% in 2021 [1], and sector-specific settings (e.g., assisted living communities) have also seen rapid growth in EHR and health information exchange (HIE) functions [2]. Pre-pandemic assessments document a decade of progress (2009–2019), but also reveal gaps in interoperability and use that impact team coordination [3].

Concurrently, artificial intelligence (AI)—including large language models (LLMs), retrieval-augmented generation (RAG), and multimodal learning—is increasingly applied to clinical decision support (CDS). Recent reviews highlight both emerging utility and persistent adoption barriers, including trust, explainability, and workflow fit [4, 5, 6]. Most existing systems remain narrow, supporting a single step (e.g., triage, risk prediction, imaging) rather than the longitudinal, multidisciplinary workflow clinicians require.

Context: National deployments (DoctorSV)

Recent national initiatives such as El Salvador's DoctorSV signal momentum toward unified digital health services and a "single clinical history" vision. Public communications describe phased onboarding and integration with pharmacies, laboratories, and imaging providers. While independent evaluations are pending, these programs underscore the need for modular, standards-aligned, and auditable AI frameworks that can integrate with national infrastructures [7].

2. Significance and Innovation

The project contributes to medical informatics by:

- Integrating multimodal AI components across screening-to-follow-up within one longitudinal workflow.
- Embedding transparency, provenance, and human oversight into each AI interaction (audit trails, citations, model cards).
- Providing a plug-in architecture compatible with FHIR and open data models, enabling site-specific extensibility.
- Addressing trust, usability, and task–technology fit from the outset through mixed-methods evaluation with clinicians.

Implications for national platforms

Although our evaluation uses synthetic cases in a research prototype, the framework's emphasis on FHIR compatibility, provenance/audit logs, and modular AI suggests a natural path for integration with emerging national platforms (e.g., DoctorSV), contingent on available interfaces, governance agreements, and safety validation. We did not analyze or evaluate any national deployment; references to such programs are purely contextual.

3. Objectives and Research Questions

Primary research question: How can an AI-augmented, EHR-centered framework improve information continuity and perceived decision support across multidisciplinary care?

Secondary research questions:

1. Which design principles enable effective multi-phase AI integration?
2. How do clinicians perceive trust, explainability, usability, and workflow fit when using the framework?
3. What technical and organizational challenges arise during early adoption in simulated settings?

4. Conceptual Framework

We model the encounter as five interconnected stages:

1. **Screening & Triage:** symptom capture, patient-reported information, AI-assisted risk estimation.

2. **Diagnostic Reasoning:** multimodal data integration and differential support with literature-backed rationale.
3. **Treatment Planning:** guideline-aware plan synthesis and safety checks (e.g., interactions).
4. **Prescription & Care Execution:** structured orders and care plans documented within the unified record.
5. **Follow-up & Prognosis:** monitoring, prediction, and longitudinal review.

A unified longitudinal record maintains structured fields (labs, vitals, medications), unstructured notes, and AI-generated artifacts with provenance.

5. Methodology

5.1. Study Design

A mixed-methods, within-subjects, counterbalanced study combining iterative system development with early evaluation using synthetic cases and clinician participants.

5.2. Baseline Workflow (Standard): OpenMRS Reference Application

The baseline condition uses the OpenMRS Reference Application (RefApp) with the FHIR2 module (FHIR R4). We will load the same synthetic cases used in the HPH condition into OpenMRS via FHIR Bundle POSTs. Baseline sessions disable any decision-support or AI add-ons. Participants interact with standard OpenMRS views for: Notes, Labs, Medications, Imaging/Reports, and a patient timeline. Role-based accounts (IM/FP, EM, Psychiatry) will have read/enter permissions appropriate for data review and note entry.

5.3. HPH Workflow (Prototype)

The HPH condition uses our prototype UI backed by the identical FHIR Bundles. The prototype renders the same clinical content and adds assistive features (e.g., structured summaries, literature-backed suggestions, provenance). The prototype does not expose information unavailable in the baseline.

5.4. Parity Rules (Pre-registered)

We enforce parity to isolate the effect of the framework:

- **Seeding and interaction:** Both conditions use the identical canonical FHIR R4 Bundle per case; OpenMRS is pre-seeded via FHIR2 (pre-session), and the HPH prototype reads the same Bundle. Clinicians interact only through the respective UIs; no data are entered during sessions.
- **Content parity:** identical synthetic cases, timestamps, and missingness patterns in both conditions.
- **Task parity:** identical task lists (e.g., initial orders, risk estimate, disposition).
- **Time parity:** matched time windows per block.
- **Feature parity:** no decision-support in baseline; assistive features in HPH are labeled and auditable.

5.5. Synthetic Case Generation and Artifact Pipeline

Two EHR-formatted vignettes (acute and chronic multimorbidity) will be constructed from guidelines/textbooks and refined by 2–3 clinicians. Each case is authored as a canonical **FHIR R4 Bundle** (Patient, Encounter, Observation for labs/vitals, MedicationStatement/Request, DiagnosticReport for ECG/CXR text, ImagingStudy metadata, AllergyIntolerance, CarePlan, Composition). From the Bundle we derive UI artifacts for both conditions:

- **OpenMRS baseline:** create records via FHIR POSTs; use native OpenMRS views for review.
- **HPH prototype:** read the same Bundle to render equivalent views plus assistive summaries.

Each case folder contains a manifest (hashes, counts of resources) and frozen version to ensure reproducibility.

5.6. Participants

We will recruit 6–12 physicians across specialties (e.g., internal medicine, family medicine, emergency, psychiatry). Inclusion criteria: licensed MD, $\geq$ 1 year of routine EHR use. Exclusion: prior contribution to this study's prototype design beyond informal feedback.

5.7. Session Flow and Counterbalancing

Each participant completes both cases under both workflows. Order is counterbalanced across participants (Latin square: Case A/B $\times$ Baseline/HPH). A typical session (40–60 minutes): consent and orientation; Block 1 (case $\times$ workflow); Block 2 (the alternate workflow on the same case); repeat for the second case; questionnaires and short interview.

5.8. Measures

Quantitative (5-point Likert): perceived usefulness, decision-support quality, information continuity, explainability, workflow fit; optional NASA-TLX workload. **Qualitative:** interview themes on strengths, limitations, collaborative use, and safety.

5.9. Analysis

Wilcoxon signed-rank tests (and effect sizes) compare framework-assisted versus baseline workflows within-subjects. Thematic analysis (inductive coding) is applied to interview transcripts. Mixed-methods triangulation integrates quantitative and qualitative findings.

6. Ethical Considerations

No real patient data will be collected or analyzed. Vignettes are synthetic and de-identified. Clinicians (the only human participants) will provide informed consent. The protocol will be submitted to an institutional ethics committee to confirm compliance with regulations for studies involving clinicians and simulated cases.

7. Risk Management

Risks include data misuse, overreliance on AI suggestions, and participant fatigue. Mitigations: strict separation from real EHRs; clear labeling of AI outputs as assistive; capped session length; and audit trails for all AI artifacts.

8. Timeline

Phase	**Duration & Key Tasks**
Design & Requirements	Months 1–3: literature review; architecture; IRB prep.
Prototype Development	Months 4–6: build plugins; unify record; interface.
Evaluation	Months 7–8: clinician sessions; data collection.
Analysis & Writing	Months 9–10: analysis; manuscript; preregistration.

9. Dissemination Plan

We will prepare a preprint and target a digital health venue (e.g., JMIR AI, JAMIA Open, Frontiers in Digital Health). Open-source components that do not process patient data will be released under a permissive license, with documentation and demo notebooks.

10. Code and Data Availability

All non-proprietary software components and study protocols will be released under a permissive open-source license (e.g., BSD-3-Clause). A code availability statement will be included as recommended by JMIR [8], with repository URL and an archived DOI (immutable release) cited in the manuscript.

11. Expected Outcomes

- A validated conceptual and technical framework for AI-integrated encounters.
- A working prototype demonstrating cross-phase AI interoperability.
- Empirical insights into clinician trust, explainability, and workflow integration in simulated studies.

12. References

References

[1] Centers for Disease Control and Prevention. National electronic health records survey (nehrs): Results and publications (2021 cycle). https://www.cdc.gov/nchs/nehrs/results/index.html, May 2024. Accessed November 2025.

[2] Shih-Chuan Lin and Özge Tunalılar. Rapid adoption of electronic health record and health information exchange among assisted living communities, 2010–2018. *Journal of the American Medical Informatics Association*, 29(5):953–957, 2022.

[3] Jiexi X. Jiang et al. A decade of progress in electronic health record adoption from 2009 to 2019: a pre-pandemic assessment. *Health Services Research*, 2023. Epub ahead of print.

[4] H. M. Tun et al. Trust in artificial intelligence–based clinical decision support systems: Systematic review. *Journal of Medical Internet Research*, 27:e69678, 2025.

[5] Hugo M. Vasquez, Enrico Pianarosa, Radu Sîrbu, et al. Human factors methods in the design of digital decision support systems for population health: a scoping review. *BMC Public Health*, 24:2458, 2024.

[6] B. D. Simon et al. The future of multimodal artificial intelligence models for healthcare. *npj Digital Medicine*, 8, 2025.

[7] CAF – Development Bank of Latin America and the Caribbean. El salvador launches doctorsv, a digital health system pioneer in the world. https://www.caf.com/en/currently/news/el-salvador-launches-doctorsv-a-digital-health-system-pioneer-in-the-world/, November 2025. Accessed November 2025.

[8] JMIR Publications. Jmir data and code availability policy. https://support.jmir.org/hc/en-us/articles/360030832631-What-is-the-JMIR-Publications-data-sharing-policy, 2025. Accessed November 2025.

Appendices

A. Synthetic Vignette and Administration Protocol

Vignette A (Acute Care: Chest Pain in Primary/Urgent Setting)

Chief complaint: "Chest pain for 3 hours."

Triage note (nurse, arrival T0):

- 54-year-old male; abrupt retrosternal pressure while lifting boxes.
- Pain 7/10, radiates to left arm; associated diaphoresis and mild nausea.
- No syncope; no dyspnea at rest. Walk-in; arrived by car.

Past medical history:

- Hypertension (10 years), hyperlipidemia, former smoker (20 pack-years).
- Father died of MI at 58.

Medications/allergies:

- Amlodipine 10 mg daily; Atorvastatin 40 mg nightly.
- No known drug allergies.

Vital signs at T0:

- BP 156/92 mmHg; HR 104 bpm; RR 18; Temp 36.8°C; SpO_2 97% RA.

Initial tests (available at T+15 min):

- ECG: sinus tachycardia; non-specific ST-T changes (no ST elevation).
- Troponin I (high-sensitivity): pending at T+15; available at T+60.
- Basic labs (CBC, BMP, lipid panel): pending.

Imaging (available on request):

- Portable chest X-ray: not yet performed.

Progress note (physician, T+60):

- Troponin I (hs): 24 ng/L (ULN 18); repeat at T+120 pending.
- Pain decreased to 3/10 after rest; persists with exertion.

Uncertainty/missingness intentionally included:

- No prior ECG on file; medication adherence unclear.
- Exact symptom onset time self-reported; no external records available.

Artifacts to load in the prototype:

- Triage note, vital trend table, structured meds list, ECG text report, labs table (with timestamps), imaging placeholder, timeline view.

Role-based prompts (to display in-session)

- **Generalist (IM/FP):** Identify immediate risks, initial orders, and disposition. How does the framework support information continuity?
- **Emergency:** Triage safety, time-sensitive decisions, and escalation criteria. What signals are surfaced at a glance?
- **Psychiatry/Mental Health:** Screen for anxiety/panic overlay, medication interactions, and follow-up continuity. Which longitudinal elements matter to you here?

Administration protocol (within-subjects, counterbalanced)

1. **Design:** 2 (workflow: standard vs. prototype) $\times$ 2 (case: A vs. B) within-subjects; Latin-square counterbalancing of order.
2. **Flow per participant (40–60 min):**
   (a) Consent; 2-min orientation; start think-aloud.
   (b) Block 1: Case (A or B) with *standard workflow*. (10–12 min)
   (c) Block 2: Case (A or B) with *prototype*. (10–12 min)
   (d) Repeat Blocks 1–2 with the other case, swapping workflow order.
   (e) Post-task questionnaires (Likert scales) and 8–10 min interview.
3. **Measures:** usefulness, decision support, explainability, information continuity, workflow fit, optional NASA-TLX.
4. **Artifacts captured:** screen/audio recording, timestamps, orders placed, notes generated (de-identified), questionnaire responses.

Sample size rationale

With a within-subjects design, each clinician experiences both workflows on both cases, allowing paired comparisons. Usability and early mixed-methods studies commonly use $n = 6$–12 to identify the majority of major issues and reach thematic saturation across roles while keeping sessions feasible. Stratifying specialties (e.g., IM/FP, EM, Psychiatry) ensures role-specific feedback.

Notes

Case B (chronic, multimorbidity) is prepared similarly (e.g., type 2 diabetes, hypertension, depression, polypharmacy) with longitudinal labs, meds changes, and missingness (e.g., gaps in adherence). This appendix illustrates structure and administration; all values are synthetic

and for research use.

B. Physician Interaction Protocol

Purpose

This appendix specifies exactly how clinicians interact with the baseline (OpenMRS) and the HPH prototype to ensure parity, reproducibility, and clarity for replication.

Systems under comparison

- **Baseline (Standard):** OpenMRS Reference Application with FHIR2 (FHIR R4). No decision-support/AI modules enabled.
- **HPH (Prototype):** HPH UI reading the identical FHIR Bundle per case. Assistive features (summaries, rationale, provenance) are enabled and labeled as such.

Data seeding and interaction rule

Both conditions use the *identical* canonical FHIR R4 Bundle per case; OpenMRS is pre-seeded via FHIR2 *before* sessions, and HPH reads the same Bundle. Participants interact only through the respective UIs. No manual data entry or edits are performed during sessions.

Pre-session setup (facilitator checklist)

1. Verify the correct role-based account (IM/FP, EM, Psychiatry) is logged in for the participant.
2. Open the assigned case (*A* or *B*) on the landing page of the designated system (Baseline or HPH) according to the counterbalanced order sheet.
3. Confirm that Notes, Labs, Medications, Imaging/Reports, and Timeline views are accessible.
4. Start screen and audio capture; confirm consent and think-aloud.
5. Set a visible countdown timer for the block (10–12 minutes).

Session design and timing

Within-subjects, counterbalanced (Latin square: Case A/B × Baseline/HPH). Each participant completes **both** cases in **both** workflows.

1. Block 1: Case (A or B) with *Workflow 1* (Baseline or HPH) (10–12 min).
2. Block 2: Same Case with *Workflow 2* (HPH or Baseline) (10–12 min).
3. Repeat Blocks 1–2 with the other Case (B or A).

4. Post-task questionnaires (Likert scales) and 8–10 min interview.

Task checklist (display to participants)

- Identify immediate risks and red flags.
- Propose initial orders/actions (labs, imaging, meds) and disposition.
- Provide a short written rationale (2–3 lines) in the notes area.
- Identify information gaps or missing data that could change decisions.

Role prompts (shown on a card beside the screen)

- **IM/FP:** Coherent plan for first visit/urgent setting; continuity considerations for follow-up.
- **EM:** Time-sensitive safety and escalation; admission vs. DC.
- **Psych:** Anxiety/panic overlay, interactions, and longitudinal risks; how continuity artifacts assist judgment.

Baseline (OpenMRS) navigation map

1. **Landing:** Patient dashboard for the assigned case.
2. **Notes:** Read triage and progress notes; add a brief rationale (if a notes area is provided for the study user).
3. **Labs:** Review time-stamped labs and reference ranges.
4. **Medications:** Review current/previous meds and adherence notes.
5. **Imaging/Reports:** Read ECG/CXR text reports (no images).
6. **Timeline:** Cross-check event order (onset → tests).

HPH navigation map

1. **Overview:** Read structured summary and provenance panel.
2. **Notes:** Read triage/progress excerpts with citations.
3. **Labs:** Review trends/flags (values equal to Baseline).
4. **Medications:** Review list; see interaction checks if present.
5. **Reports:** Read ECG/CXR text; follow provided citations.
6. **Timeline:** Verify event order and latency between steps.

Interaction constraints (parity enforcement)

- No external search, no copy/paste into other tools.
- No new orders actually placed; all actions are hypothetical.
- No editing of patient data; notes are study artifacts only.
- Stay within the allocated time; notify when done.

Captured artifacts and measures

- Screen/audio recording with timestamps.
- Notes/rationales entered (study area only).
- Task completion indicators (checklist).
- Likert outcomes: usefulness, decision support, explainability, information continuity, workflow fit; optional NASA-TLX.

Moderator prompts (use sparingly)

- "Please think aloud as you work."
- "What would you do next, and why?"
- "What information is missing or uncertain?"
- "How did this view help or hinder your decision?"

Deviations and fail-safes

- If a page fails to load, pause the timer, refresh once, then continue.
- If a role account logs out, resume with the same case and workflow.
- Record any deviation and its duration in the session log.