CASE STUDY · HEALTHCARE AI

AI-Driven General Physician Pod for Precision Health Guidance

A multi-prompt conversational diagnostic assistant that turns unstructured patient descriptions into structured, clinically-grounded guidance.

GP Pod is a patient-facing virtual health assistant. Patients describe symptoms in plain language; the system collects the key clinical variables it is missing, reasons across a chain of specialised prompts, and returns a triage recommendation with clear next steps — while explicitly deferring to licensed clinicians for diagnosis and treatment.

Domain: Digital Health / Clinical Decision Support | **Solution Type:** Multi-Prompt LLM Chatbot
Deliverable: Patient Diagnostic & Triage Platform | **Document:** Detailed Case Study

1. Executive Summary

Primary care systems are saturated. A large share of general-physician consultations are low-acuity questions that could be resolved with reliable guidance, while genuinely urgent cases sit in the same queue. At the same time, patients searching symptoms online receive unranked, context-free information that produces either false reassurance or unnecessary alarm.

The General Physician Pod (GP Pod) addresses this gap. It is a multi-prompt chatbot that acts as a first-line health companion: it interviews the patient for the information a clinician would ask for, structures that information, reasons over it through a sequence of specialised prompts, and returns a precision health recommendation — self-care, scheduled consultation, or urgent escalation — together with the reasoning behind it.

3–5×	**<60s**	**24/7**	**100%**
Faster intake vs. manual forms	Median time to guidance	Availability, no queue	Sessions logged & auditable

2. The Problem

2.1 For Patients

- **Information without context.** Search engines return the same results for a 25-year-old with a mild headache and a 68-year-old with a sudden severe one.
- **Access friction.** Appointment waits, cost, and geography delay questions that need answering the same day.
- **Poor self-triage.** Patients routinely under-react to red-flag symptoms and over-react to benign ones.

2.2 For Providers

- Clinician time is consumed by intake questions that are highly repetitive and structurable.
- Consultations begin with incomplete histories, adding minutes per patient.
- Low-acuity walk-ins crowd capacity needed for higher-acuity cases.

2.3 Why a Single-Prompt Chatbot Fails Here

A generic single-prompt LLM assistant tends to answer immediately from incomplete input, blends safety-critical triage with general explanation, cannot be audited step-by-step, and has no structural point at which a safety rule can be enforced. Clinical guidance requires the opposite: bounded steps, explicit state, and a hard safety gate.

3. The Solution: GP Pod

GP Pod decomposes the consultation into discrete, independently governed reasoning stages. Each stage has one job, one prompt, and one validated output contract. The output of each stage becomes structured input to the next, so the system never asks a downstream model to do work it has no information for.

> **Design principle:** the chatbot never guesses at missing clinical variables. If a required field is absent, the system asks for it before proceeding to reasoning.

3.1 The Multi-Prompt Chain

Stage	Responsibility	Output
1. Intake	Parses free-text complaint; extracts symptom, onset, duration, severity, location, character.	Structured symptom object
2. Safety Gate	Screens for emergency red flags before any other reasoning runs. Runs on every turn, not just the first.	Escalate / Continue
3. Clarification	Identifies missing required fields and generates the minimum set of follow-up	Question queue

	questions.	
4. Context Assembly	Merges answers with profile data: age, sex, comorbidities, medications, allergies, pregnancy status.	Complete patient context
5. Differential Reasoning	Generates ranked candidate explanations with supporting and contradicting evidence from the context.	Ranked hypothesis set
6. Triage Classification	Maps hypotheses and severity onto an urgency tier.	Self-care / Routine / Urgent / Emergency
7. Guidance Composer	Writes the patient-facing answer at plain-language reading level, with rationale and next steps.	Final response
8. Compliance Review	Verifies no diagnosis is asserted, no prescription is issued, disclaimer is present, and tier matches content.	Approved / Rewrite

3.2 Key Capabilities

- **Guided information capture** — the patient answers a short, adaptive set of questions instead of filling a static form.
- **Context-aware precision** — the same symptom yields different guidance depending on age, comorbidity, and medication profile.
- **Always-on red-flag detection** — emergency screening runs independently of the conversational flow and can interrupt it at any turn.
- **Transparent reasoning** — every recommendation is delivered with the factors that drove it.
- **Continuity** — session memory allows follow-up questions without re-entering history.
- **Clinician handoff** — the structured context object exports as a pre-consultation summary.

4. Architecture

Layer	Function
Conversational Interface	Web and mobile chat surface; accessible, multilingual-ready, mobile-first.
Orchestration Engine	Executes the prompt chain, manages state transitions, enforces stage ordering, handles retries.
Prompt Library	Versioned, individually testable prompt templates — one per reasoning stage.
Clinical Knowledge Layer	Retrieval over curated, source-attributed medical references to ground responses and reduce fabrication.
Safety & Guardrail Layer	Deterministic red-flag rules, scope boundaries, prohibited-output filters, mandatory disclaimers.
Patient Context Store	Encrypted profile and session history; supports longitudinal follow-up.
Audit & Analytics	Full stage-by-stage trace of every session for clinical review, quality assurance, and regulatory evidence.

5. Illustrative Session

The following walkthrough demonstrates the chain in operation. It is an illustration of system behaviour, not a clinical record.

Stage	Behaviour
Patient input	"I've had a headache for three days and it's worse in the morning."
Intake	Extracts: symptom = headache; duration = 3 days; pattern = morning-predominant. Flags missing: severity, age, medication history, associated symptoms.

Safety Gate	No emergency red flags present in the initial statement. Continue.
Clarification	Asks four targeted questions covering severity, onset character, associated symptoms, and current medications.
Context Assembly	Merges the answers with the stored profile to produce a complete context object.
Differential Reasoning	Produces a ranked hypothesis set with the evidence supporting and contradicting each.
Triage	Assigns an urgency tier and the timeframe for acting on it.
Guidance	Returns plain-language explanation, the reasoning behind the tier, self-care measures where appropriate, and the specific symptoms that should trigger immediate escalation.
Compliance	Confirms no diagnosis asserted, no medication prescribed, disclaimer attached. Approved.

6. Safety, Ethics and Governance

6.1 Hard Boundaries

- GP Pod does not diagnose. It provides guidance and triage.
- GP Pod does not prescribe, adjust, or discontinue medication.
- GP Pod does not interpret imaging or laboratory results as a substitute for a clinician.
- GP Pod does not manage psychiatric crisis; such conversations route immediately to human crisis support.

6.2 Escalation-First Bias

Where evidence is ambiguous, the system is deliberately tuned to escalate rather than reassure. Under-triage carries asymmetric harm compared with over-triage, and the safety gate is weighted accordingly.

6.3 Governance Controls

- Clinical review board approves every prompt version before release.
- Red-flag rules are deterministic code, not model judgement.
- All sessions are logged with full reasoning traces for retrospective audit.
- Data handling follows healthcare privacy requirements: encryption at rest and in transit, role-based access, explicit consent, and defined retention limits.
- Sampled sessions undergo ongoing clinician review; disagreements feed prompt revision.

Medical disclaimer. GP Pod is an informational and triage support tool. It is not a medical device, does not provide a diagnosis, and does not replace consultation with a qualified healthcare professional. In an emergency, patients should contact local emergency services immediately.

7. Measured Outcomes

Dimension	Before	With GP Pod
Patient information capture	Static forms, frequently incomplete	Adaptive questioning, required fields enforced
Time to initial guidance	Hours to days	Under a minute
Consistency of advice	Variable by source	Deterministic prompt chain, versioned
Red-flag detection	Dependent on patient judgement	Systematic, every turn
Clinician intake burden	Full history taken in-session	Structured summary available pre-consultation
Auditability	None	Complete stage-level trace

7.1 Metrics Tracked

- **Triage concordance** — agreement between GP Pod's urgency tier and an independent clinician's tier on the same case.
- **Under-triage rate** — the primary safety metric; monitored continuously with alerting thresholds.
- **Question efficiency** — number of clarifying turns required to reach a complete context.
- **Completion rate** — proportion of sessions reaching a guidance output.
- **Patient comprehension** — post-session confirmation that the recommendation and next steps were understood.

8. Implementation Approach

Phase	Focus	Activities
1. Discovery	Scope definition	Select target condition set, define red-flag taxonomy with clinical partners, establish triage tiers and success criteria.
2. Design	Prompt architecture	Author each stage prompt, define output schemas, build the deterministic guardrail rule set.
3. Build	System assembly	Orchestration engine, knowledge retrieval layer, context store, conversational interface, audit pipeline.
4. Clinical Validation	Safety proof	Run curated case sets against clinician-assigned ground truth; tune until under-triage falls below threshold.
5. Pilot	Controlled release	Limited cohort with mandatory clinician review of every escalation; collect concordance data.
6. Scale	Production operation	Expand condition coverage, add languages, integrate with scheduling and EHR handoff, continuous monitoring.

9. Why the Multi-Prompt Design Matters

- **Testability.** Each stage can be evaluated in isolation against its own ground truth, so a regression is traceable to one prompt.
- **Controllability.** Safety logic lives in a dedicated gate rather than being diluted across one large instruction block.
- **Explainability.** The reasoning trace is a genuine artefact of execution, not a post-hoc narrative.
- **Maintainability.** Clinical guidance updates touch a single prompt, not the whole system.
- **Cost efficiency.** Simple stages route to smaller models; only complex reasoning consumes frontier capacity.

10. Conclusion

GP Pod demonstrates that clinical-grade conversational AI is an architecture problem before it is a model problem. By decomposing a consultation into governed stages — intake, safety screening, clarification, context assembly, reasoning, triage, composition, and compliance review — the system delivers guidance that is contextual, consistent, transparent, and safe by construction.

The result is a virtual health assistant that gives patients a reliable first answer at the moment they need it, and gives clinicians a structured, higher-quality starting point when the patient reaches them.

> **Next step.** Adapt the GP Pod architecture to a specific condition set, patient population, or care-delivery model, with clinical validation designed against your own governance requirements.