# **“Mindful Diagnostic Tool” with bounded agents, MCP tools, and dual-knowledge memory**

## **1) Control plane (genome-driven)**

* **Digital Genome (v1)** — YAML/JSON spec of services, workflows, policies, safety rules, SLOs, rollback playbooks; signed & versioned.
* **Genome Runner (autopoiesis)** — interprets the genome to deploy/repair/migrate components (K8s operators or equivalents), emits lifecycle events, and enforces identity during migrations. [MDPI](https://www.mdpi.com/2073-431X/13/9/220?utm_source=chatgpt.com)
* **Software Workflow Manager** — executes clinical workflows from the genome (triage → evidence gathering → hypothesis generation → explanation → human gate).
* **Cognizing Oracles** — policy/ethics/teleonomy checks; confidence & contradiction analysis; explanation with provenance; all ML/LLM outputs pass through here before action. [MDPI](https://www.mdpi.com/2078-2489/13/1/24?utm_source=chatgpt.com)

## **2) Knowledge & memory**

* **Shared Knowledge Graph (Neo4j)** — canonical medical graph (Patient/Encounter/Observation/Medication/Procedure/Allergy/Imaging, plus disease/syndrome relations, UMLS/SNOMED/ICD mappings).
* **Individual (Local) KGs** — per service/agent: role-specific patterns, execution logs, decision histories; tuned for fast, local retrieval.
* **Event Store (append-only)** — Kafka → Lake (Iceberg/Delta) capturing **observation→inference→action** chains; is the **ground truth** for reflection, audit, and safety cases. [Preprints](https://www.preprints.org/manuscript/202406.1622/v1?utm_source=chatgpt.com)
* **Reflector service** — computes drift, quality deltas, and safety metrics; proposes genome/policy updates backed by event evidence; humans approve changes.

## **3) Agentic layer (bounded)**

* **MCP Tool Adapters** (inside app):  
  + *Data curation agent* — validates & normalizes clinical texts/images before KG ingestion.
  + *Evidence summarizer agent* — drafts patient-specific evidence summaries & counterfactuals; Oracle checks and clinician approves.
  + *Coding/helper agents* — propose query plans, refactor pipelines, generate docs (dev environment or canary only).
* **Hard rules:** agents are **propose-only**; they cannot write to the shared KG, models, or genome directly. All proposals go through **Oracles + Reflector + human gates**.

## **4) Clinical safety & governance**

* **SaMD lifecycle hooks** — tie event metrics to release gates; model change control; post-market surveillance from the event store; human-in-the-loop approvals for high-risk outputs.
* **Teleonomy registry** — explicit goals/constraints (safety > cost > latency; therapy policies; “never suggest unapproved treatment”). Oracles enforce these at run time. [Sciety](https://sciety.org/articles/activity/10.20944/preprints202410.2564.v1?utm_source=chatgpt.com)

## **5) Security posture (zero-trust + provenance)**

* **mTLS/service mesh + OIDC/ABAC** on all APIs and graph access; PHI vault segregation; prompt/response redaction for any MCP calls; **signed genome releases**; all agent/tool calls logged to the **event store** with full provenance for replay. (This cleanly supports FDA/ISO audit trails.)

# **Where each prior artifact lands**

* **Development Crew:** becomes a **delivery accelerator** (R&D and canary), and in prod becomes **bounded assistants** behind Oracles and genome-defined workflows. Keep the orchestrator pattern; scope tools to least privilege; require proposal → review → apply.
* **Dual-Knowledge:** becomes your **memory fabric**: local KGs for specialization + a governed shared KG, synchronized by **event-triggered** promotions with abstraction levels. This implements **associative memory** and a clean **meta-cognitive loop**. [ResearchGate](https://www.researchgate.net/publication/381721520_Digital_Genome_and_Self-Regulating_Distributed_Software_Applications_with_Associative_Memory_and_Event-Driven_History?utm_source=chatgpt.com)

# **Guardrails & failure-mode notes (important)**

* **Drift between local and shared KGs** → mitigate with scheduled summarization, schema validation, and human-approved promotions.
* **Agent/tool jailbreak or prompt leakage** → constrain MCP surfaces (allow-lists), constant PHI redaction, time-boxed execution, and policy engines in front of tools.
* **Over-automation risk** → “no-write” contract for agents; only genome runner mutates runtime state after Oracle + human approval. [MDPI](https://www.mdpi.com/2073-431X/13/9/220?utm_source=chatgpt.com)

# **Step-by-step rollout (8–10 weeks to MVP)**

1. **Data & memory plane**: stand up shared KG (with core clinical ontologies) + event store; add minimal local KG schema for 2 services (ingestion, summarizer).
2. **Genome v1 + Runner**: encode 3 workflows (triage, evidence gathering, rare-disease flag); wire self-healing and migration basics. [Preprints](https://www.preprints.org/manuscript/202406.1622/v1?utm_source=chatgpt.com)
3. **Oracles**: implement policy/ethics/confidence checks and explanation with provenance paths from KG + event chains. [MDPI](https://www.mdpi.com/2078-2489/13/1/24?utm_source=chatgpt.com)
4. **Bounded agents via MCP**: add two tool adapters (curation, evidence summarization); enforce proposal-only path through Oracles.
5. **Dual-Knowledge sync**: implement event-triggered promote/demote flows and abstraction levels; add reviewer UI for merges.
6. **Safety gates**: wire SaMD change-control & human-in-the-loop approvals to the Reflector metrics.

\