MemoriAI – High‑Level Design (HLD) – Advanced

Version 2.0 • Owner: Group 3 (Ayvazoglu, Bovilla, Kondeti, Sejwal, Singh)  
Date: 2025-10-15

# 1. Executive Summary

MemoriAI is an AI‑powered digital companion for Alzheimer’s care with three MVP pillars: Cognitive & Identity Assist, Daily Reminders & Safety, and Caregiver Dashboard. The design favors REST APIs, privacy by design, and guardrailed LLM usage.

# 2. Goals & Non‑Goals

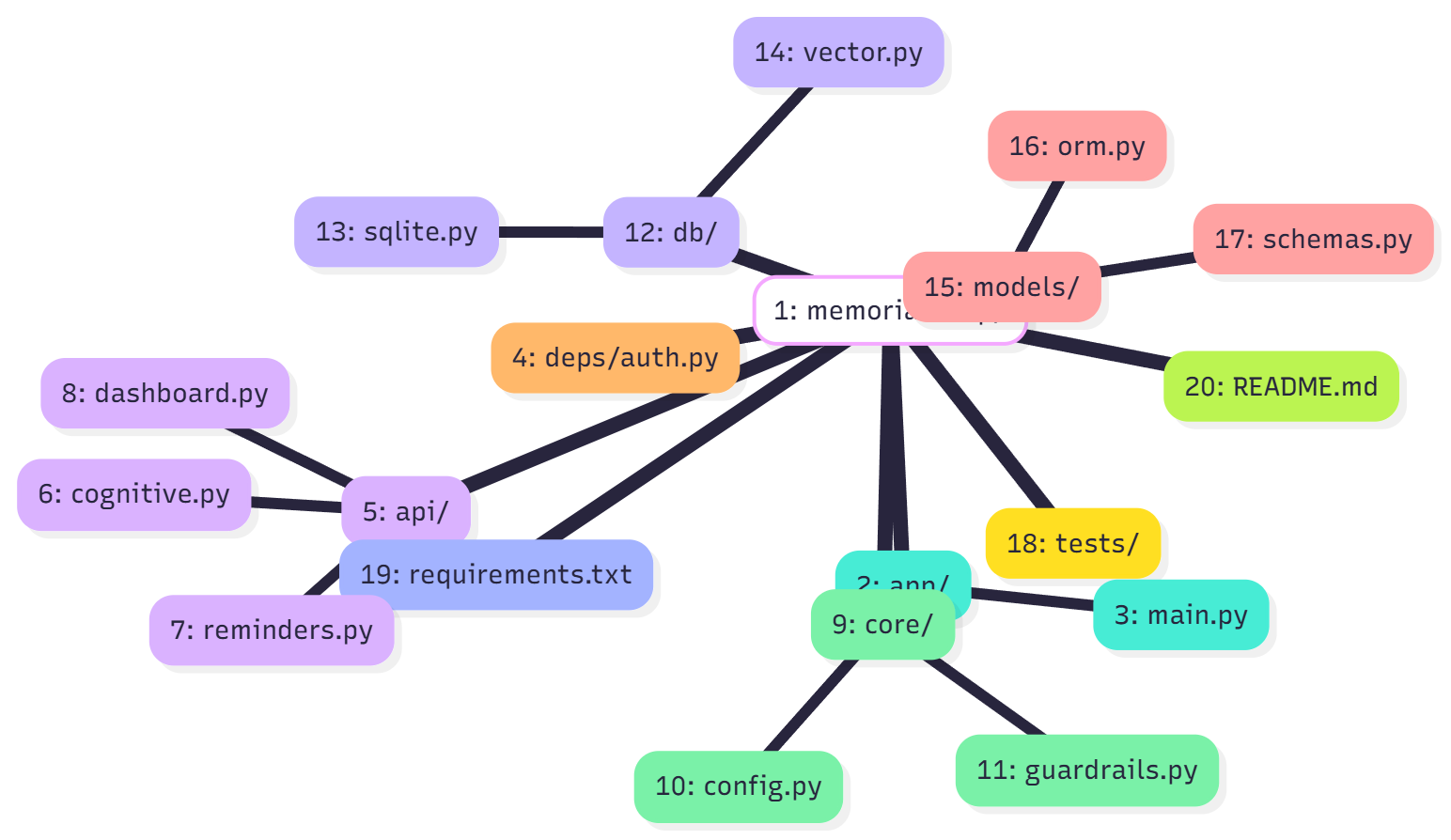
|  |  |
| --- | --- |
| Goals | Non‑Goals |
| Facts recall, adaptive reminders, caregiver summaries. | Diagnosis or clinical decision support. |
| Guardrailed LLM, RESTful APIs, privacy‑first data. | Training a foundation model or multi‑tenant SaaS. |

# 3. Personas & Use Cases

P1 Patient • P2 Caregiver • P3 Clinician (read‑only) • P4 Admin

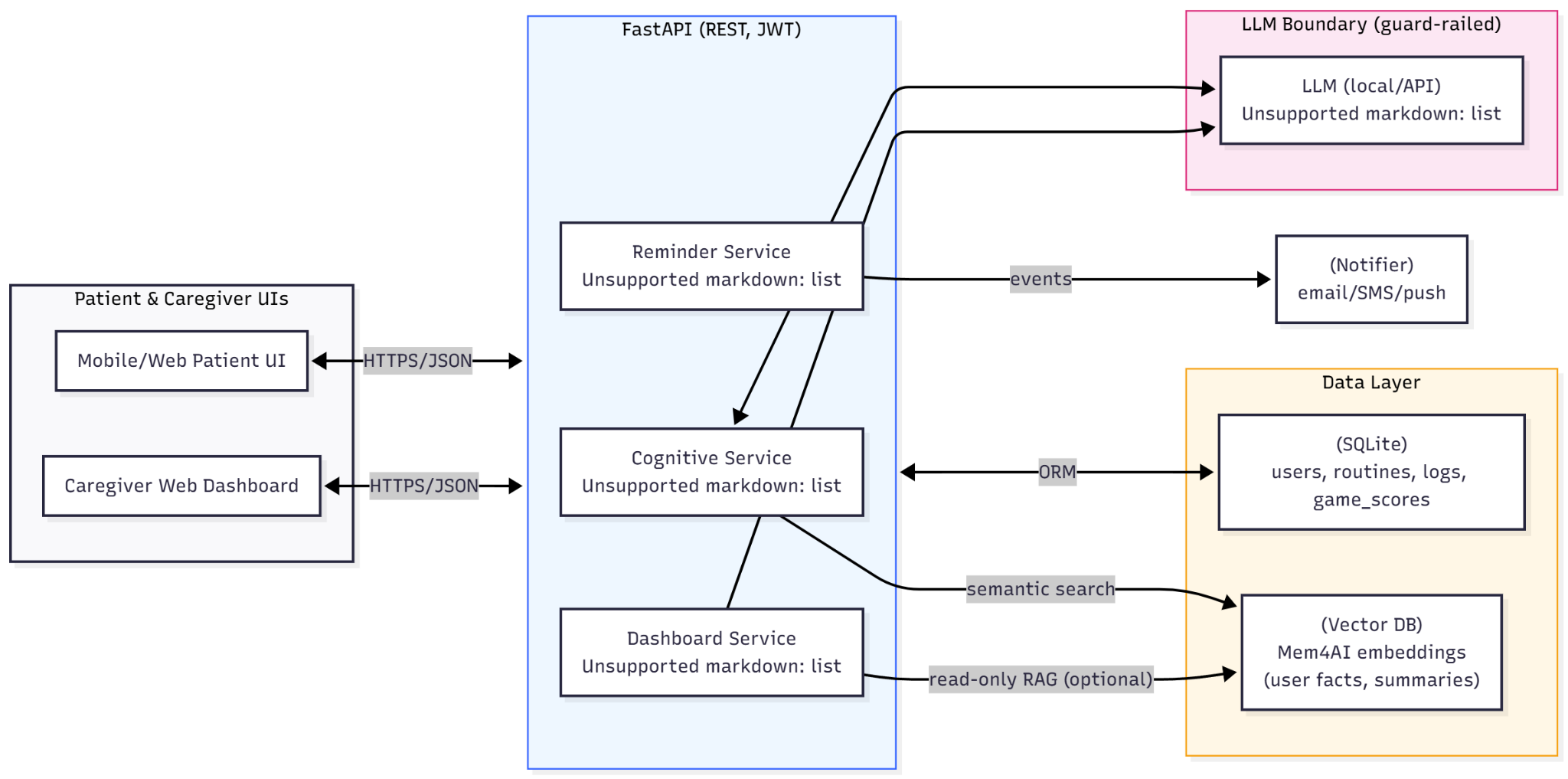
UC‑A Cognitive & Identity Assist; UC‑B Daily Reminders; UC‑C Caregiver Dashboard.

# 4. System Context



# 5. Architecture Overview

FastAPI services: Cognitive, Reminders, Dashboard. Data: SQLite + Vector DB (Mem4AI). Guarded LLM for phrasing only; identity recall is facts‑only.



# 6. Component Responsibilities

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Component | Responsibilities | Tech | Dependencies | Owner |
| Cognitive Service | facts recall, word‑aid, semantic search | FastAPI, LangChain | SQLite, VectorDB, LLM | AI Lead |
| Reminder Service | scheduling, notifications, adherence | FastAPI, APScheduler | SQLite, Notifier | Backend Dev |
| Dashboard Service | summaries, metrics | FastAPI, Pandas | SQLite (read) | Full‑stack |
| Auth/RBAC | JWT/OAuth2, roles | fastapi-security, pyjwt | SQLite | Backend Dev |
| Notifier | email/SMS push | SMTP/Twilio (mock ok) | Reminder | Full‑stack |

# 7. Public API Contracts (MVP)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Service | HTTP | Route | Request→Response | Notes |
| Cognitive | POST | /cognitive/ask | {"q":"Who is Anna?"} → {"answer":"Anna is your granddaughter."} | facts only |
| Cognitive | POST | /cognitive/word-aid | {"prefix":"ila"} → {"suggestions":[...]} | semantic top‑K |
| Reminders | POST | /reminders | {"time":"14:00","text":"Take pills"} → 201 | create |
| Reminders | POST | /reminders/mark-done | {"id":123} → 200 | adherence |
| Dashboard | GET | /dashboard/summary?date=YYYY‑MM‑DD | → {"adherence":0.82,...} | no LLM |

# 8. Data Model & Storage

|  |  |  |  |
| --- | --- | --- | --- |
| Entity | Fields | Purpose | PII |
| users | id, role, name, email\_hash, created\_at | auth & ownership | low |
| routines | id, user\_id, time, text, status | reminders | no |
| adherence\_logs | id, routine\_id, ts, status | tracking | no |
| game\_scores | id, user\_id, game, score, ts | trends | no |
| mem4ai\_embeddings | vec\_id, user\_id, embedding, metadata | semantic memory | pseudo |

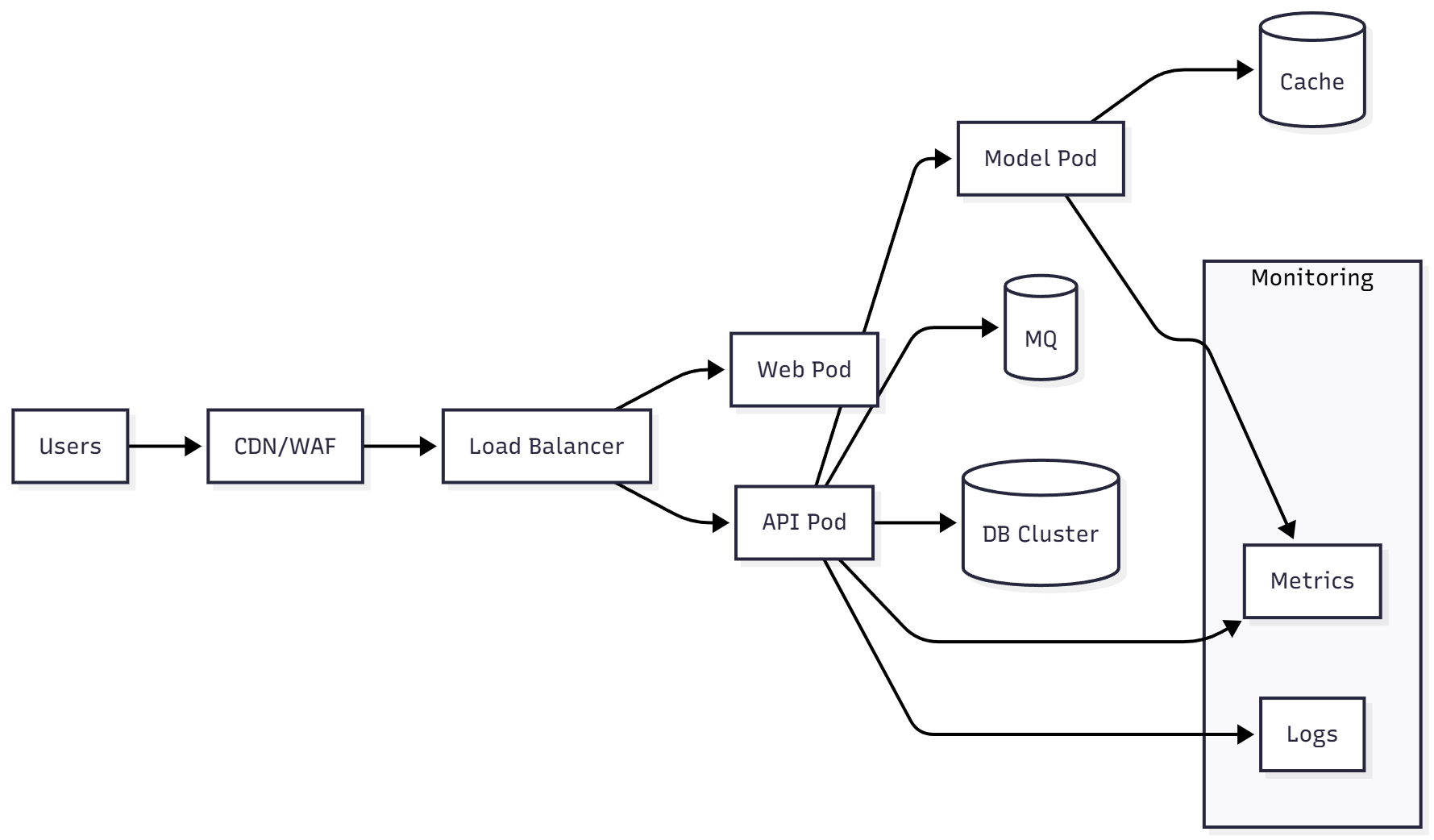
# 9. Security, Privacy & Safety

|  |  |  |
| --- | --- | --- |
| Control | Implementation | Notes |
| AuthN/Z | OAuth2 + JWT; RBAC | Short‑lived tokens, refresh flow |
| Transport | TLS 1.2+ | HSTS / CSRF protections |
| Data at rest | SQLite encryption; secrets in env | Key rotation, least privilege |
| PII minimization | Only essential facts; user‑editable memory | Right‑to‑erase |
| AI guardrails | System prompts + refusal policies | No medical advice |
| Audit | Structured logs + access trails | ELK/CloudWatch when hosted |

# 10. Reliability (SLOs) & Performance

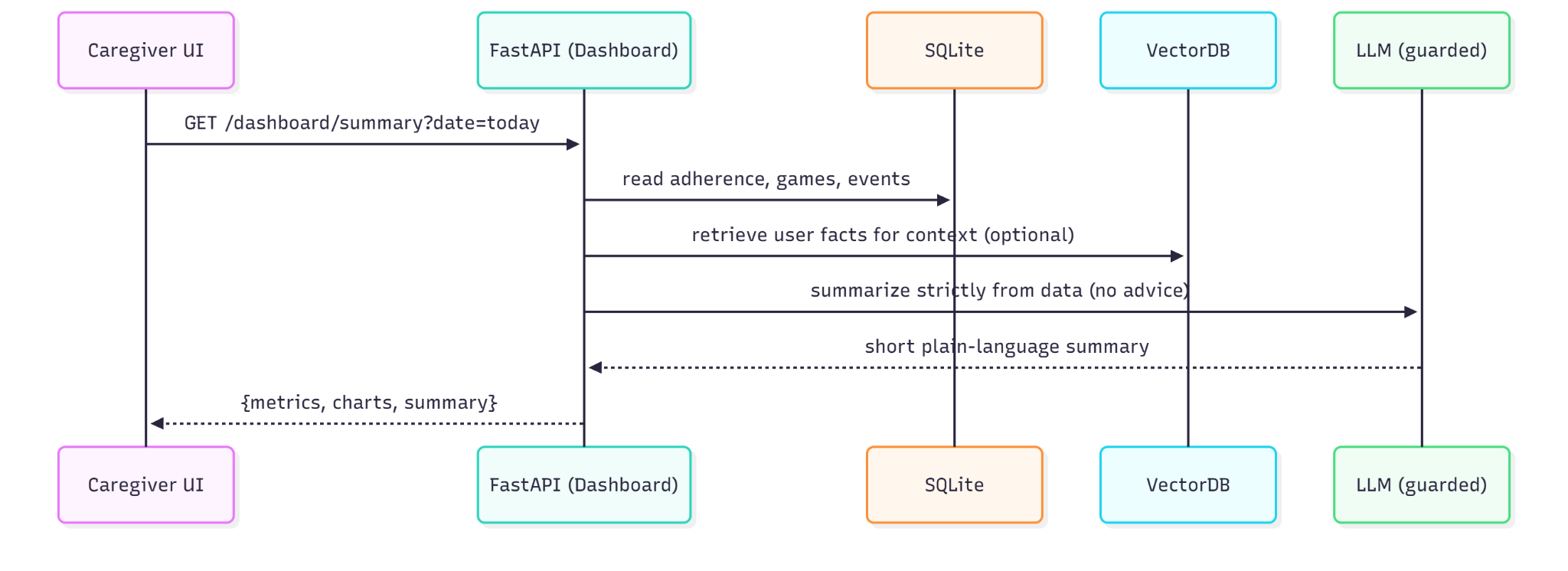
|  |  |  |
| --- | --- | --- |
| Metric | Target | Validation |
| p95 latency | <1.5s | k6 load test |
| Uptime | 99% | health checks |
| Reminder on‑time | >98% | synthetic tests |

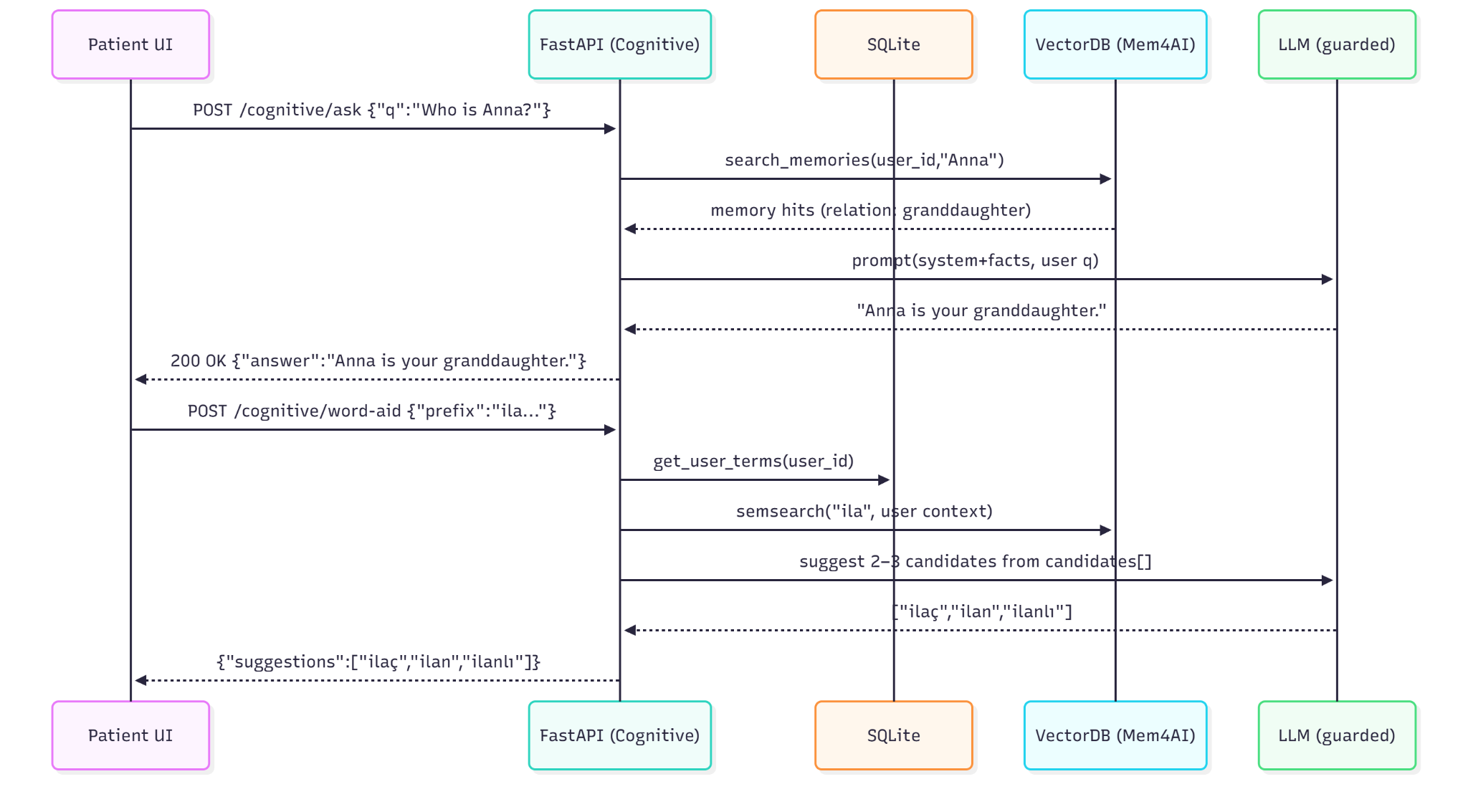
# 11. CI/CD & Environments

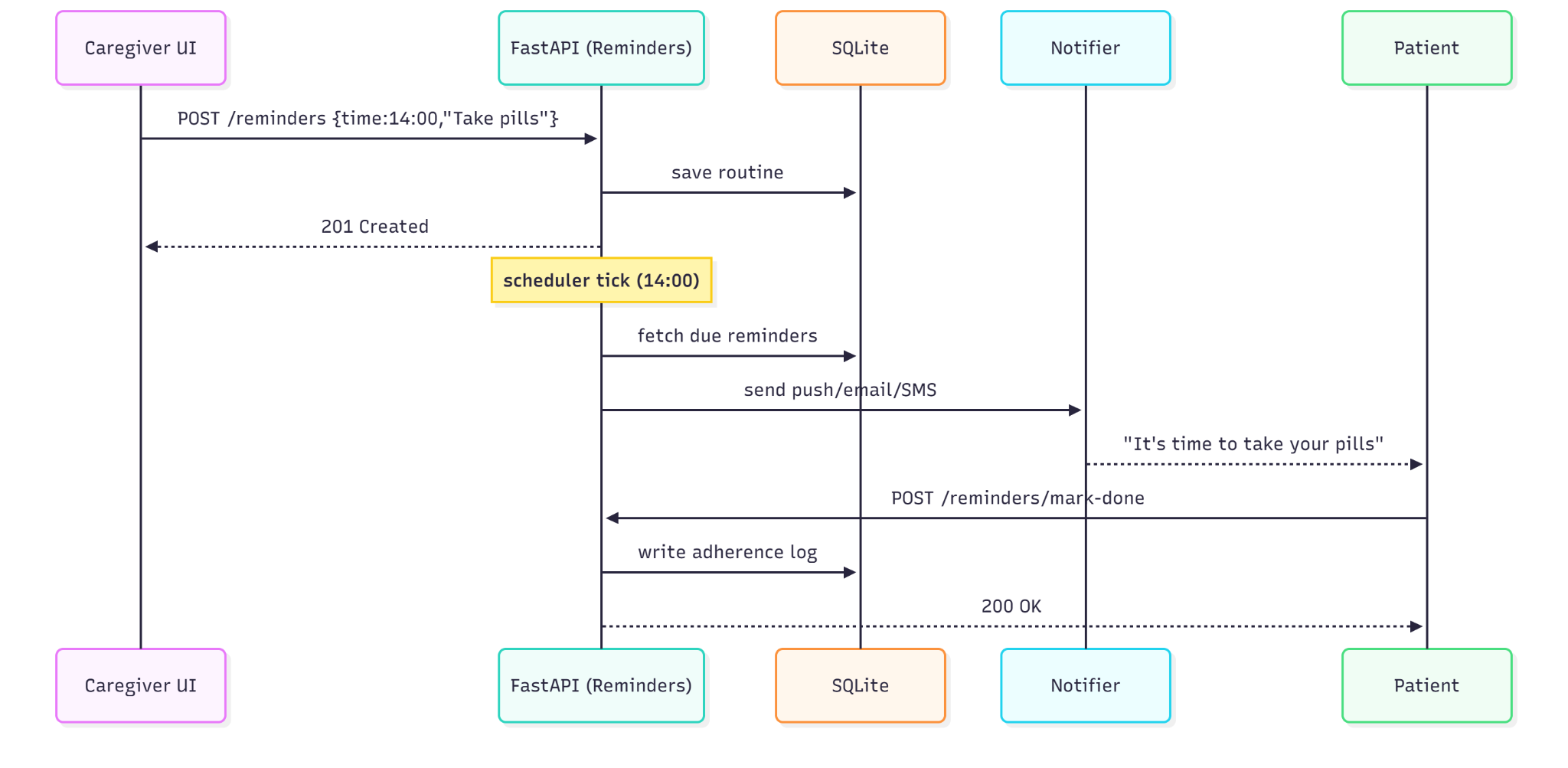


CI: GitHub Actions (lint, unit tests, security checks). CD: Docker build/push; compose/k8s manifests.

# 12. Key Flows







# 13. Detailed ERD & DDL

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Entity | Column | Type | Constraint | Index |
| users | id | INTEGER | PK AUTOINCREMENT | PK\_users |
| users | role | TEXT | NOT NULL CHECK(role IN ("patient","caregiver","admin")) | idx\_users\_role |
| users | name | TEXT | NOT NULL | idx\_users\_name |
| users | email\_hash | TEXT | UNIQUE | idx\_users\_emailhash |
| users | created\_at | DATETIME | DEFAULT CURRENT\_TIMESTAMP |  |
| routines | id | INTEGER | PK AUTOINCREMENT | PK\_routines |
| routines | user\_id | INTEGER | FK -> users(id) | idx\_routines\_user |
| routines | time | TEXT | NOT NULL (HH:MM) | idx\_routines\_time |
| routines | text | TEXT | NOT NULL |  |
| routines | status | TEXT | CHECK(status IN ("active","paused")) | idx\_routines\_status |
| adherence\_logs | id | INTEGER | PK AUTOINCREMENT | PK\_adherence |
| adherence\_logs | routine\_id | INTEGER | FK -> routines(id) | idx\_adherence\_routine |
| adherence\_logs | ts | DATETIME | NOT NULL | idx\_adherence\_ts |
| adherence\_logs | status | TEXT | CHECK(status IN ("done","missed")) | idx\_adherence\_status |
| game\_scores | id | INTEGER | PK AUTOINCREMENT | PK\_games |
| game\_scores | user\_id | INTEGER | FK -> users(id) | idx\_games\_user |
| game\_scores | game | TEXT | NOT NULL | idx\_games\_game |
| game\_scores | score | INTEGER | NOT NULL |  |
| game\_scores | ts | DATETIME | NOT NULL | idx\_games\_ts |
| mem4ai\_embeddings | vec\_id | TEXT | PK | PK\_mem4ai |
| mem4ai\_embeddings | user\_id | INTEGER | FK -> users(id) | idx\_mem4ai\_user |
| mem4ai\_embeddings | embedding | BLOB | NOT NULL |  |
| mem4ai\_embeddings | metadata | TEXT | JSON string | idx\_mem4ai\_meta |

SQLite DDL (excerpt) provided in Appendix A.

# 14. JWT Claims & RBAC Matrix

|  |  |  |
| --- | --- | --- |
| Claim | Type/Example | Notes |
| iss | https://auth.memoriai.local | Issuer |
| sub | user:123 | Subject (user id) |
| aud | memoriai-api | Audience |
| iat | 1697350000 | Issued at (epoch) |
| exp | 1697357200 | Expiry (epoch) |
| role | patient|caregiver|admin | RBAC role |
| scope | reminders:write reminders:read cognitive:read | Fine-grained scopes |
| ver | 1 | Token schema version |

RBAC Matrix:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Role | Cognitive | Reminders | Dashboard | Admin |
| patient | ASK/READ | READ + ACK only | SELF metrics only | No |
| caregiver | ASK/READ | CRUD for assigned patient | FULL caregiver view | No |
| admin | READ logs | READ | READ | Yes: user mgmt, configs |

# 15. Metrics & Dashboards

|  |  |  |  |
| --- | --- | --- | --- |
| Panel | Metric | Source | Goal/Alert |
| API Health | p95 latency, error\_rate | Prometheus + FastAPI middleware | Alert if p95 > 1.5s (5m) |
| Reminder Pipeline | sent, delivered, ack rate | SQLite + notifier logs | Ack ≥ 80%, delivery ≥ 98% |
| Cognitive Usage | ask\_count, word\_aid\_accept\_rate | App logs | Word-aid accept ≥ 70% |
| LLM Safety | refusal\_count, hallucination\_suspects | Guardrail logs | Zero medical advice |
| DB Health | DB size, slow queries | SQLite metrics | < 200ms query p95 |

# 16. Capacity Planning (Rough Sizing)

Assumptions: 100 DAU, 10 req/day → ~1,000 RPD. Burst: 5 RPS (1 min). Single node (2 vCPU, 4 GB RAM) sufficient. Scale plan: Postgres + hosted vector store + API replicas.

|  |  |  |  |
| --- | --- | --- | --- |
| Resource | Baseline | Burst Target | Notes |
| CPU | < 40% | < 70% | Limited LLM calls; cache |
| Memory | < 1.5 GB | < 3 GB | Small vector index |
| Disk | < 1 GB | < 5 GB | SQLite + embeddings |
| Network | < 5 Mbps | < 20 Mbps | Notifications + API |

# 17. Risk Matrix

Likelihood (L) and Impact (I) scored 1–5. Priority = L × I.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risk | L | I | Priority | Mitigation | Owner |
| Hallucination in answers | 2 | 5 | 10 | Facts-only recall; guardrails; manual tests | AI Lead |
| Reminder delivery failures | 3 | 4 | 12 | Retries; ack flow; synthetic monitors | Backend Dev |
| PII leakage | 2 | 5 | 10 | Minimization; encryption; RBAC; audits | Security |
| Scope creep | 4 | 3 | 12 | MVP gates; change control; sprint planning | PM |
| Single-node failure | 2 | 4 | 8 | Backups; health checks; restart policy | DevOps |

Flow Captions (Actor Clarification):

A) Caregiver Dashboard — Primary actor: \*\*Caregiver\*\*. System aggregates adherence and interactions; \*\*Patient\*\* is secondary (data source).

B) Cognitive & Identity / Word Aid — Primary actor: \*\*Patient\*\*. LLM is guardrailed; facts-only recall; caregiver is not required.

C) Daily Reminder Flow — Primary actor: \*\*Patient\*\* (receives notification and acknowledges). \*\*Caregiver\*\* configures reminders (secondary).