AI Concierge

Doc type: technical · **Version:** v0.1 · **Status:** published · **Module slug:** ai-concierge
Exported: 2026-05-15 11:11 UTC · **By:** anonymous

AI Concierge – Technical Specification

1. Module Purpose & Scope (Authoritative)

AI Concierge is Primoro's voice and call-orchestration module. It handles inbound calls during out-of-hours and overflow periods, executes governed operational outbound recovery calls for short-notice cancellations, and provides operational support for lead intake, triage assistance, and consultation booking when clinical teams are busy or unavailable. Its purpose is to reduce missed calls and hold times, handle routine requests safely and consistently, and convert all call activity into trackable work.

It governs:

- Inbound call handling in Out-of-Hours and Overflow modes, including intent detection, multi-turn dialogue, and identity verification.
- Rota-triggered outbound recovery for short-notice appointment cancellations, with controlled retries and suppression lifecycle management.
- Workflow automation from call content: creating call threads, generating tasks, sending forms, and triggering approved follow-ups via Communication Hub and Task Manager.

It explicitly does not:

- Own or mutate Lead/Opportunity pipeline stage, BookingEligibility, SLA, or next-action — those are owned by Treatment Pipeline Manager.
- Own or modify appointment records — those are owned by Appointment Manager.
- Own marketing campaign enrolment, segmentation, or marketing performance attribution — those are owned by Campaign Manager.
- Provide clinical diagnosis or treatment advice.
- Mislead callers about being human.

2. Ownership & Responsibilities

2.1 AI Concierge IS Responsible For

- Inbound call handling in Out-of-Hours and Overflow modes.
- Intent detection and multi-turn dialogue for supported self-service actions.
- Real-time transcription with timestamps and speaker labels, stored immutably in call threads.
- Caller identification, disambiguation, and identity verification prior to any sensitive disclosure.
- Policy recital using approved wording (deposits, assessment-only visits, cancellation windows).
- Staff Agent Console integration inside Communication Hub for takeover and one-click actions (RBAC-governed).

- Workflow automation from call content: sending forms, creating tasks, and triggering approved follow-ups.
- Telephony integration (launch platform: 3CX) and routing maps for departments and queues.
- Suppression lifecycle signalling to Campaign Manager during and after active concierge recovery workflows.
- Full auditability of all call events, verification events, AI actions, and staff actions.

2.2 AI Concierge IS NOT Responsible For

- Lead/Opportunity creation governance, stack/stage management, BookingEligibility, or SLA/next-action — owned by Treatment Pipeline Manager.
- Appointment record modification or booking authority — owned by Appointment Manager.
- Marketing campaign initiation, opt-in journey management, or marketing attribution — owned by Campaign Manager.
- Clinical diagnosis, treatment advice, or clinical note access.
- People scoring, ranking, or performance policing.
- Acting as a passive surveillance product.

3. Core Objects (Normative)

3.1 Call Thread (Canonical Artefact)

A Call Thread is a governed digital artefact representing the complete, immutable record of a single call interaction.

Ownership boundary: AI Concierge is the originating author of Call Thread data — it creates the thread, appends transcript segments, records detected intents, and manages state transitions. Communication Hub owns the Call Thread as the canonical container within its data model; AI Concierge writes to and reads from the Call Thread via Communication Hub's call thread management API (see §6.3). The two modules share the same underlying artefact: AI Concierge does not maintain a parallel shadow copy. Any reference to "Call Thread" in this specification refers to the Communication Hub-hosted canonical record authored and managed by AI Concierge through the declared integration contract.

Minimum required fields:

- CallThreadID
- FK to patient / contact record
- CallMode (Out-of-Hours / Overflow / Recovery)
- CallState (see §5.1)
- CreatedAt (timestamp)
- StartTime / EndTime / Duration
- DetectedIntents
- FinalOutcome
- VerificationState (see §5.2)
- AuditTrail (immutable)

3.2 Call Thread State Machine (Authoritative)

States:

- Ringing
- AI Answered
- Identifying Caller
- Verifying Identity
- Handling Intent
- Handover Requested
- In Staff Takeover
- Completed
- Follow-Up Created

Rules:

- State transitions are auditable and time-stamped.
- Calls MUST NOT end without a recorded outcome.
- Unresolved calls MUST transition to Follow-Up Created and generate a Task Manager task before closing.
- Calls MAY NOT return to Ringing or AI Answered once in Handover Requested or later states.
- Handover Requested may be initiated by either the caller request or AI-side escalation logic; in both cases the transition MUST be logged with the initiating actor.

3.3 Identity Verification State Machine (Authoritative)

States:

- Unverified
- Verification Prompted
- Verified (method recorded)
- Verification Failed

Rules:

- Patient-specific details MUST NOT be disclosed prior to Verified.
- Shared numbers MUST trigger "Who is this regarding?" disambiguation where family linkage exists.
- Verification Failed limits AI responses to generic information only; AI MUST offer handover to staff.
- The verification method used (e.g. date-of-birth challenge, known-caller match) MUST be recorded in the Call Thread's AuditTrail at the point of transition to Verified.

4. Operational Capabilities

4.1 Inbound Call Handling (Authoritative)

The module MUST:

- Answer all inbound calls in configured Out-of-Hours and Overflow modes.
- Introduce itself transparently as an AI assistant; it MUST NOT represent itself as human.
- Detect caller intent and engage multi-turn dialogue to service supported self-service actions.

- Complete caller identification and, where sensitive data is involved, identity verification before disclosure.
- Recite policy wording (deposits, assessment-only visits, cancellation windows) from the approved wording library only.
- Create or attach to exactly one Call Thread per call.
- Produce a recorded outcome on every call without exception.

The module MAY:

- Answer read-only patient record questions (upcoming appointments, outstanding forms/tasks, unpaid balances, whitelisted demographic fields) after successful verification.
- Confirm upcoming appointment dates, times, locations, and clinicians from Appointment Manager (read-only).
- Capture reschedule intent and surface governed options without committing changes.
- Send confirmation messages via Communication Hub where configured.

The module MUST NOT:

- Modify appointment records — all changes require staff handover or a governed booking UI in Appointment Manager.
- Disclose clinical notes, diagnoses, or any data absent explicit verification.
- Alter patient records.

4.2 Short-Notice Appointment Recovery — Rota-Triggered Outbound (Authoritative)

This is an operational workflow, not a marketing campaign. All activity is excluded from Campaign Manager statistics.

Trigger: Appointment Manager emits a `RotaCancellationEvent` to AI Concierge when an appointment is cancelled and the appointment start time falls within the configurable short-notice window (e.g. same-day or next-day). See §6.2 for the full inbound event contract. AI Concierge MUST NOT initiate outbound recovery independently of this event.

Rota availability constraints: Before initiating or continuing outbound recovery attempts, AI Concierge SHOULD consult Rota Manager to confirm that at least one member of staff capable of accepting a recovery booking is available within the recovery window. If Rota Manager indicates that no suitable staff availability exists for the relevant slot (e.g. all clinicians are absent or the schedule pattern has no remaining capacity), AI Concierge MUST NOT initiate outbound recovery calls for that slot and MUST instead create a Task Manager follow-up for human intervention. This prevents contacting patients about slots that cannot be filled. Where real-time Rota Manager availability cannot be determined (e.g. integration unavailability), AI Concierge MUST fall back to creating a Task Manager follow-up rather than proceeding blindly.

The module MUST:

- Subscribe to rota cancellation events from Appointment Manager.
- Assess eligibility for recovery using: appointment type and duration; patient suitability rules; consent and contact preferences; contact-fatigue and suppression rules.
- Execute outbound recovery calls via the telephony integration with controlled retries and configurable back-off.
- Capture responses, intent, and outcomes in a Call Thread.
- On acceptance: escalate to staff or the governed booking flow per Appointment Manager rules.

- On non-acceptance or timeout: continue recovery attempts up to configured thresholds.
- On exhaustion: create a Task Manager follow-up for human intervention.
- Emit suppression signals to Campaign Manager per the lifecycle defined in §6.1.

The module MUST NOT:

- Bypass Appointment Manager booking governance.
- Enrol contacts into Campaign Manager marketing journeys.
- Continue outbound attempts beyond configured retry thresholds without human review.

4.3 Lead Intake — Out-of-Hours and Overflow (Governed)

When a call is received outside staffed hours or when queues exceed thresholds:

The module MUST:

- Identify whether the caller presents a new enquiry.
- If a new enquiry: create or link to exactly one Lead via Treatment Pipeline Manager intake endpoints, with no duplication.
- Instantiate the Lead into the appropriate configured Stack (e.g. New Patients, Emergency) per Treatment Pipeline Manager rules.
- Populate the Lead record with the fields required by Treatment Pipeline Manager's governed intake schema at the point of creation. Required fields include: `SourceChannel` (set to `ai-concierge`), `TriageStatus` (set to the outcome of any triage questions asked during the call), and `ConversationThreadID` (set to the CallThreadID of the originating call). AI Concierge MUST NOT set `BookingEligibility` to `true` at intake — it MUST be initialised to `false` and progressed only by Treatment Pipeline Manager's own eligibility logic.

The module MAY:

- Ask triage-relevant questions defined by the active Stack.
- Reference required Digital Triage flows without completing them.
- Summarise triage context into the Lead or Opportunity record.

The module MUST NOT:

- Determine BookingEligibility.
- Skip required triage steps.
- Mutate pipeline stage or BookingEligibility directly.

4.4 Consultation Booking Support (Governed)

The sole authority for booking permission is `Lead.BookingEligibility = true` or `Opportunity.BookingEligibility = true`, as owned by Treatment Pipeline Manager and enforced by Appointment Manager.

The module MAY:

- Surface booking options when BookingEligibility is confirmed true.
- Hand over to staff or open the governed booking UI.

The module MUST NOT:

- Bypass BookingEligibility.
- Force or infer booking consent.

All booking attempts MUST be logged against the call thread and the pipeline entity.

4.5 Staff Handover (Authoritative)

States: Eligible for Handover → Routing to Queue → Staff Connected → AI Steps Out.

Rules:

- Handover MUST include the transcript, intent summary, and context card.
- Transcript MAY continue recording during the human portion where configured.
- Handover MUST be logged with actor, timestamp, and reason.

5. Delivery Surfaces & Access (Authoritative)

5.1 Staff Web Portal — Communication Hub Calls Workspace (Primary)

The Agent Console within Communication Hub MUST include:

- **Live Call Queue** — status, duration, mode, assigned queue.
- **Screen-Pop Context Card** — identity status, upcoming appointments, recent context, open forms, open tasks, family disambiguation.
- **Live Transcript Panel** — streaming transcript with key-moment markers.
- **Operator Guidance Panel** — intent-aware, policy-safe scripts.
- **One-Click Actions (RBAC-governed)** — takeover, transfer, create task, send form, send payment link (where enabled), open booking UI.

5.2 Tablet App

(no content captured in original — needs definition)

5.3 Patient Mobile App

AI Concierge does not present a direct patient mobile app surface; patient-facing outputs are spoken responses during calls and links sent via approved channels (SMS/email) through Communication Hub.

5.4 Engagement Signals

- Call thread outcomes and intent summaries surfaced to staff via Communication Hub.
- Outcome signals (reached, accepted, declined, no-answer) emitted to Campaign Manager.
- Engagement metrics — call volumes by mode, recovery acceptance rates, and handover frequency — are available for staff analytics via Communication Hub reporting.

6. Integration Contracts

6.1 Suppression Lifecycle — AI Concierge ↔ Campaign Manager (Authoritative)

Activation: Suppression is applied at the point AI Concierge begins an active concierge recovery workflow for the contact — specifically when the first outbound recovery attempt is initiated or an inbound recovery call is answered.

Lift conditions: Suppression MUST be lifted, and an explicit suppression-lifted event MUST be emitted to Campaign Manager, upon whichever of the following occurs first:

- The recovery workflow reaches a terminal outcome (accepted, declined, exhausted retries, or staff follow-up task created); or
- A configurable suppression expiry window elapses without a terminal outcome (default: 48 hours from activation; MUST be tenant-configurable).

Explicit signal: AI Concierge MUST emit a suppression-lifted event to Campaign Manager when suppression ends, regardless of reason. Silent continuation of suppression after workflow conclusion is prohibited.

Re-suppression: If a second concierge recovery workflow is initiated for the same contact before Campaign Manager has resumed enrolment, the suppression window MUST be extended from the new workflow activation point under the same rules.

Failsafe: Campaign Manager MUST NOT resume enrolment for a suppressed contact until it receives the suppression-lifted signal or the expiry window has elapsed with no signal received.

Permitted cross-links:

- AI Concierge MAY emit engagement/outcome signals to Campaign Manager (reached, accepted, declined, no-answer).
- AI Concierge MAY suppress marketing enrolment while a contact is actively engaged in a concierge recovery workflow.
- Campaign Manager MAY reference concierge-derived engagement metadata when proposing campaigns or exclusions.

Prohibited overlap:

- Campaign Manager MUST NOT initiate voice calls autonomously.
- AI Concierge MUST NOT enrol contacts into marketing campaigns.
- Rota-triggered recovery is operational, not marketing, and MUST be excluded from campaign statistics.

6.2 Inbound Integration Contracts

From module	What	Contract
Appointment Manager	Rota cancellation events (short-notice window)	Async event / webhook
Appointment Manager	Read-only appointment data (dates, times, clinicians)	Sync read
Treatment Pipeline Manager	Lead/Opportunity BookingEligibility, Stack configuration	Sync read

Treatment Pipeline Manager	Lead intake endpoints (create/link Lead)	Sync write (intake only)
Communication Hub	Call thread container creation and management	Sync
Access Manager	RBAC roles and permissions for Agent Console actions	Sync read
Rota Manager	Staff availability and schedule patterns for recovery slot validation	Sync read

RotaCancellationEvent contract (Appointment Manager → AI Concierge):

AI Concierge MUST subscribe to the `RotaCancellationEvent` emitted by Appointment Manager. The event payload MUST include, at minimum:

- `appointment_id` — the unique identifier of the cancelled appointment.
- `contact_id` — the patient/contact associated with the appointment.
- `appointment_type` — used to assess recovery eligibility per configured rules.
- `appointment_start_at` — the original scheduled start time, used to evaluate whether the cancellation falls within the configured short-notice window.
- `cancellation_reason_code` — the reason category as recorded by Appointment Manager; AI Concierge uses this to determine whether outbound recovery is appropriate.
- `slot_duration_minutes` — required to assess whether a replacement slot of the same duration is plausible.

On receipt, AI Concierge MUST evaluate eligibility before initiating any outbound recovery attempt (see §4.2). If eligibility criteria are not met, AI Concierge MUST log the received event and its non-eligibility decision in the audit trail but MUST NOT initiate any outbound call.

6.3 Outbound Integration Contracts

To module	What	Contract
Communication Hub	Call thread events, transcript storage, task/form dispatch	Event + sync write
Task Manager	Follow-up tasks for unresolved calls and exhausted recovery	Event
Campaign Manager	Suppression activation/lift signals; engagement outcome signals	Event
Appointment Manager	Booking eligibility checks; escalation to governed booking flow	Sync read + handover

Treatment Pipeline Manager	Lead/Opportunity creation and triage context updates	Sync write (intake)
Performance Dashboards	Call volume metrics, recovery rates, verification rates, handover frequency	Metric emission (see §6.7)

Campaign Manager outcome-signal payload contract:

AI Concierge emits four named outcome signals to Campaign Manager for each recovery or inbound call interaction. Campaign Manager consumes these signals read-only for audience segmentation, step suppression, and telemetry annotation. The signals and their required payload fields are:

Signal	Emitted when	Required payload fields
`contact.reached`	AI Concierge successfully connects with the contact (call answered)	`contact_id`, `call_thread_id`, `call_mode`, `reached_at` (ISO 8601 timestamp)
`contact.accepted`	Contact accepts the recovery offer or confirms an action during the call	`contact_id`, `call_thread_id`, `call_mode`, `accepted_at`, `intent_summary`
`contact.declined`	Contact explicitly declines the recovery offer or terminates without acceptance	`contact_id`, `call_thread_id`, `call_mode`, `declined_at`
`contact.no_answer`	Call attempt receives no answer within the configured ring timeout	`contact_id`, `call_thread_id`, `call_mode`, `attempted_at`, `attempt_number`

All signals MUST be emitted as discrete events at the point the outcome is determined; batching is not permitted. Each signal MUST include an `ai_origin: true` marker and an `emitted_at` timestamp so that Campaign Manager can apply provenance labelling. Signals are operational in nature and MUST NOT be used by Campaign Manager as a basis for marketing enrolment without independent consent and eligibility checks.

6.4 PMS Boundary

The PMS is the system of record for patient demographics and clinical data. AI Concierge reads whitelisted administrative fields only (via Appointment Manager or direct integration where supported). Where PMS auto-logging is supported, call outcomes are posted back; where not, AI Concierge MUST create a Task Manager fallback for staff action. AI Concierge MUST NOT write clinical data to the PMS.

6.5 Telephony Integration

Launch telephony integration: **3CX**. AI Concierge owns routing maps for departments and queues. Future telephony providers must be introduced via a configurable integration layer without requiring changes to AI Concierge's core call-handling logic.

Rate governance: All telephony provider API traffic from AI Concierge is routed through the platform's External Provider API Gateway. AI Concierge MUST declare two distinct traffic classes for rate governance purposes:

- **Interactive** — inbound call handling and real-time dialogue (high-priority, latency-sensitive). This class MUST be allocated a dedicated rate limit quota so that interactive call traffic is never degraded by outbound recovery burst activity.
- **Background** — outbound recovery call initiation and retry scheduling (standard-priority, throughput-oriented). This class operates under a separate quota and MUST NOT consume capacity reserved for the Interactive class.

In the event of a mass rota cancellation generating a spike in Background-class outbound recovery requests, the rate governance layer MUST apply backpressure to the Background queue only; Interactive inbound call handling MUST remain unaffected. AI Concierge MUST honour rate-limit backpressure signals from the gateway by queuing retry attempts within its own configured back-off schedule rather than dropping them.

6.6 Rota Manager Integration (Inbound Read)

AI Concierge consumes availability data from Rota Manager on a read-only basis to validate that outbound recovery calls are only made for slots where staff capacity exists. The integration contract is:

- **What is read:** Active schedule patterns and real-time availability for clinicians relevant to the cancelled appointment type; specifically whether the relevant clinician (or a suitable substitute) has available capacity within the recovery window.
- **Contract type:** Sync read, invoked per recovery eligibility assessment (§4.2).
- **Failure mode:** If Rota Manager is unavailable or returns an indeterminate response, AI Concierge MUST treat availability as unconfirmed and MUST NOT proceed with outbound recovery; a Task Manager follow-up MUST be created for human review.
- **Scope:** AI Concierge reads availability state only; it MUST NOT write to, modify, or signal Rota Manager. Rota Manager is not a consumer of AI Concierge events.

6.7 Performance Dashboards — Metric Emission Contract

AI Concierge emits operational metrics to Performance Dashboards for display in platform-wide reporting. All emitted metrics MUST conform to the following provenance requirements to support Performance Dashboards' labelling standards:

- **`ai_origin` marker:** Every metric event MUST carry `"source": "ai-concierge"` and `"ai_origin": true` so that Performance Dashboards can correctly attribute and label AI-generated data.
- **Freshness timestamp:** Every metric event MUST include an `emitted_at` ISO 8601 timestamp representing the point at which the metric was calculated or the event occurred. Performance Dashboards uses this field to determine data freshness and display staleness indicators.
- **Metric catalogue:** The following metrics are emitted by AI Concierge to Performance Dashboards:

Metric	Description	Granularity
`call_volume_by_mode`	Count of calls by CallMode (Out-of-Hours, Overflow, Recovery)	Per tenant, per period
`call_volume_by_outcome`	Count of calls by FinalOutcome (resolved, handed-over, follow-up-created)	Per tenant, per period

`recovery_acceptance_rate`	Proportion of recovery calls resulting in `contact.accepted` outcome	Per tenant, per period
`recovery_exhaustion_rate`	Proportion of recovery workflows that reached retry exhaustion	Per tenant, per period
`verification_success_rate`	Proportion of calls reaching Verified state vs. Verification Failed	Per tenant, per period
`handover_frequency_by_type`	Count of handovers by initiating actor (caller-requested vs. AI-escalated)	Per tenant, per period

Metrics are emitted on call thread closure (for per-call metrics) and on a periodic rollup schedule (for aggregate rates). Metrics MUST NOT include patient-identifiable data; all emissions are aggregate or anonymised counts.

7. AI Boundaries (Non-Negotiable)

AI MAY:

- Handle multi-turn dialogue for supported self-service intents (appointment confirmation, read-only record enquiries, policy recital, lead intake).
- Summarise call intent, transcript, and context for human staff review.
- Surface booking options when BookingEligibility is confirmed true by Treatment Pipeline Manager.
- Suggest approved follow-up actions (forms, tasks, payment links) for one-click staff approval via the Agent Console.
- Ask triage-relevant questions from approved Stack definitions.
- Execute outbound recovery calls under rota-triggered governance.

AI MAY NOT:

- Determine, override, or infer BookingEligibility.
- Modify appointment records, patient records, or pipeline stage.
- Disclose patient-specific information prior to confirmed identity verification.
- Provide clinical diagnosis, clinical advice, or access clinical notes.
- Enrol contacts into marketing campaigns.
- Represent itself as human.
- Bypass RBAC, audit, or governance checks.
- Make commitments on behalf of the practice outside governed self-service boundaries.

8. Audit & Compliance

The system MUST log:

- Call and Call Thread identifiers, mode, timestamps, and duration.

- Every Call Thread state transition with actor (AI, patient/caller, staff) and timestamp.
- Detected intents and final outcome state on every call.
- All identity verification events, including verification method and outcome.
- All AI-generated responses and suggestions, flagging which were accepted, overridden, or ignored by staff.
- All one-click Agent Console actions (takeover, transfer, task creation, form/payment dispatch) with actor and timestamp.
- All transcript access events (read, export) with actor, role, and timestamp.
- All suppression activation, extension, and lift signals emitted to Campaign Manager.
- All cross-module events consumed (rota cancellation, BookingEligibility reads) and emitted (follow-up tasks, suppression signals).
- PMS auto-log attempts and fallback task creation events.
- All `RotaCancellationEvent` receipts from Appointment Manager, including eligibility assessment outcomes (eligible / ineligible / unavailable-staff).
- All Rota Manager availability queries made during recovery eligibility assessment, including the result and any fallback actions taken.

Audit logs MUST be immutable and exportable for inspection. Transcripts and metadata MUST be encrypted at rest and in transit, and RBAC-controlled.

9. Access Control

Access control is governed by Access Manager using role-based access control (RBAC). The following capability map applies:

Capability	Roles permitted
View live call queue and transcript	Reception, Clinical Coordinator, Practice Manager
Perform Agent Console one-click actions (takeover, transfer, task, form, payment link)	Reception, Clinical Coordinator, Practice Manager
Open governed booking UI from Agent Console	Reception, Clinical Coordinator, Practice Manager
Export transcripts and audit logs	Practice Manager, Compliance Officer
Configure recovery thresholds, suppression windows, short-notice windows	Practice Admin, Practice Manager
Configure approved policy wording library	Practice Manager

The role names above are illustrative placeholders aligned with Access Manager's RBAC model; canonical role definitions are owned by Access Manager.

Agent Console actions MUST never bypass RBAC enforcement. MFA requirements for sensitive operations (e.g. transcript export, configuration changes) are governed by Access Manager policy.

10. Integration Summary

- **Communication Hub** — primary call thread container (canonical owner); outbound transcript, task, and form events; Agent Console host surface. AI Concierge creates and manages Call Thread content via Communication Hub's call thread management API.
- **Task Manager** — receives follow-up task creation events for unresolved calls and exhausted recovery workflows.
- **Appointment Manager** — source of `RotaCancellationEvent` (recovery trigger); read-only appointment data; booking governance and escalation target.
- **Treatment Pipeline Manager** — Lead/Opportunity intake target (with required fields `SourceChannel`, `TriageStatus`, `ConversationThreadID`; `BookingEligibility` initialised to `false`); source of BookingEligibility and Stack configuration; triage context target.
- **Campaign Manager** — receives suppression activation/lift signals and four named engagement outcome signals (`contact.reached`, `contact.accepted`, `contact.declined`, `contact.no_answer`); MUST NOT initiate voice calls.
- **Rota Manager** — read-only source of staff availability used to validate recovery slot feasibility before outbound calls are initiated.
- **Access Manager** — RBAC enforcement for all Agent Console actions and data access.
- **External Provider API Gateway** — rate governance for telephony provider traffic, with Interactive and Background traffic classes isolated to prevent recovery bursts from degrading inbound call handling.
- **Performance Dashboards** — receives operational metrics with `ai_origin` and `emitted_at` provenance markers.
- **Audit & Compliance** — immutable event log consumer for all call, verification, AI, and staff action events.

11. Explicit Non-Goals

- **Clinical diagnosis or advice** — outside scope for any module in the current platform; no ownership assignment.
- **Marketing outbound calling** — owned by Campaign Manager; AI Concierge MUST NOT initiate marketing calls.
- **Autonomous appointment booking without BookingEligibility** — owned and enforced by Appointment Manager and Treatment Pipeline Manager.
- **Patient record modification** — write access to patient records is not in scope for this module.
- **Passive surveillance** — call transcription and recording are strictly for operational and governance purposes only; use outside this purpose is prohibited.

12. Versioning & Governance

This specification is owned by: the AI Concierge module owner.

Changes to this spec require:

- Review by the Post-MVP module owner.
- Impact analysis across declared related modules (see /propose), particularly Communication Hub, Campaign Manager, Treatment Pipeline Manager, and Appointment Manager given the authoritative

integration boundaries defined in §3 and §6.

- Version bump (patch for clarifications; minor for capability additions; major for ownership boundary changes or breaking integration contract changes).

13. Build Contract (Engineering & QA)

13.1 Canonical Data Model

```
call_threads (
 id                  UUID PRIMARY KEY,
 contact_id          UUID NOT NULL,           -- FK to patient/contact record
 call_mode           TEXT NOT NULL,           -- Out-of-Hours | Overflow | Recovery
 call_state          TEXT NOT NULL,           -- see §5.1 state machine
 verification_state  TEXT NOT NULL,           -- Unverified | Verification Prompted | Verified | Veri
 verification_method TEXT,                    -- populated on transition to Verified
 detected_intents    JSONB,
 final_outcome       TEXT,                    -- resolved | handed-over | follow-up-created
 started_at          TIMESTAMPTZ NOT NULL,
 ended_at            TIMESTAMPTZ,
 duration_seconds    INTEGER,
 transcript_ref      UUID,                    -- FK to immutable transcript store
 linked_task_ids     UUID[],
 linked_lead_id      UUID,                    -- FK to Treatment Pipeline Manager Lead
 linked_opportunity_id UUID,
 audit_trail         JSONB NOT NULL,          -- immutable, append-only
 created_by          TEXT NOT NULL,           -- 'ai-concierge' | staff user ID
 created_at          TIMESTAMPTZ NOT NULL DEFAULT now()
)

call_transcripts (
 id                  UUID PRIMARY KEY,
 call_thread_id      UUID NOT NULL REFERENCES call_threads(id),
 content             JSONB NOT NULL,          -- immutable, timestamped, speaker-labelled segments
 is_editable         BOOLEAN NOT NULL DEFAULT false,
 staff_notes         TEXT,                    -- append-only staff annotations
 created_at          TIMESTAMPTZ NOT NULL DEFAULT now()
)

suppression_records (
 id                  UUID PRIMARY KEY,
 contact_id          UUID NOT NULL,
 call_thread_id      UUID NOT NULL REFERENCES call_threads(id),
 activated_at        TIMESTAMPTZ NOT NULL,
 expiry_at           TIMESTAMPTZ NOT NULL,    -- activated_at + tenant-configured window (default 48h
 lifted_at           TIMESTAMPTZ,
 lift_reason         TEXT,                    -- terminal-outcome | expiry | re-suppression-extension
 signal_emitted_at   TIMESTAMPTZ,
 created_at          TIMESTAMPTZ NOT NULL DEFAULT now()
)
```

13.2 Core Behaviour Rules

1. Every inbound call in Out-of-Hours or Overflow mode MUST create or attach to exactly one Call Thread before any AI response is delivered.

2. No patient-specific data MUST be disclosed in any AI response until the Call Thread's verification_state is `Verified`.

3. Shared numbers MUST trigger caller disambiguation before verification proceeds where family linkage exists in the patient record.

4. Every call MUST produce a non-null final_outcome before the Call Thread transitions to `Completed` or `Follow-Up Created`.

5. Any call reaching `Follow-Up Created` state MUST emit a task creation event to Task Manager with the call thread ID and intent summary.

6. AI Concierge MUST NOT write to `Lead.pipeline_stage`, `Lead.BookingEligibility`, `Opportunity.pipeline_stage`, or `Opportunity.BookingEligibility`; these fields are owned by Treatment Pipeline Manager.

7. AI Concierge MUST NOT write to any appointment record; all appointment modifications require handover to Appointment Manager's governed booking flow.

8. Outbound recovery calls MUST only be initiated following a `RotaCancellationEvent` from Appointment Manager; AI Concierge MUST NOT initiate outbound recovery independently.

9. Recovery outbound attempts MUST NOT exceed the tenant-configured retry threshold; exhaustion MUST trigger a Task Manager follow-up.

10. A suppression_record MUST be created at recovery workflow activation and a suppression-lifted event MUST be emitted to Campaign Manager at terminal outcome or expiry — whichever occurs first. Silent suppression continuation is prohibited.

11. AI Concierge MUST introduce itself as an AI on every call before any substantive dialogue.

12. All Agent Console one-click actions MUST be validated against Access Manager RBAC before execution; no action may bypass this check.

13. Call Transcripts MUST be immutable; the `is_editable` flag MUST always be `false`; staff annotations are append-only to `staff_notes` only.

14. All audit events MUST be appended to the call_thread's `audit_trail` JSONB and simultaneously emitted to the Audit & Compliance module.

15. When creating a Lead via Treatment Pipeline Manager's intake endpoint, AI Concierge MUST populate `SourceChannel` (value: `ai-concierge`), `TriageStatus`, and `ConversationThreadID`; it MUST NOT set `BookingEligibility` to `true`.

16. All four Campaign Manager outcome signals (`contact.reached`, `contact.accepted`, `contact.declined`, `contact.no_answer`) MUST be emitted individually at the point of outcome determination, with `ai_origin: true` and `emitted_at` timestamp included in each payload.

17. Telephony provider API calls MUST be routed through the External Provider API Gateway using the declared Interactive or Background traffic class as appropriate; AI Concierge MUST honour rate-limit backpressure by queuing within its configured back-off schedule.

18. Prior to initiating outbound recovery for any slot, AI Concierge MUST query Rota Manager to confirm staff availability; if unavailable or indeterminate, a Task Manager follow-up MUST be created and no outbound call initiated.

13.3 Configuration Surfaces

Setting	Scope	Configurable By

Short-notice recovery window (e.g. same-day, next-day)	Tenant	Practice Admin / Practice Manager
Suppression expiry window (default: 48 hours)	Tenant	Practice Admin / Practice Manager
Recovery retry thresholds and back-off	Tenant	Practice Admin / Practice Manager
Out-of-Hours and Overflow mode schedules	Tenant	Practice Admin / Practice Manager
Approved policy wording library	Tenant	Practice Manager
Whitelisted patient record fields for AI disclosure	Tenant	Practice Manager
Transcript recording continuation during staff takeover	Tenant	Practice Manager
Payment link enablement	Tenant	Practice Admin
Telephony routing maps (departments, queues)	Tenant	Practice Admin

13.4 Filtering & Views

The Agent Console Calls Workspace MUST support:

- Filter by call mode (Out-of-Hours, Overflow, Recovery).
- Filter by call state (active, completed, follow-up created).
- Filter by queue / department.
- Filter by date/time range.
- Filter by final outcome (resolved, handed-over, follow-up-created).
- Filter by verification state (Unverified, Verified, Verification Failed) to support compliance review.
- Saved views per staff role.

13.5 Module Extension Map

- Additional telephony providers beyond 3CX MUST be introduced via a configurable telephony integration layer without changes to the Call Thread data model or AI dialogue engine.
- Additional supported self-service intents (beyond those in §4.1) are addable without breaking the state machine; intent enumeration is additive.
- Additional one-click Agent Console actions are addable via Access Manager role mapping without schema changes.
- The suppression lifecycle (§6.1) is parameterised by tenant configuration; adding new terminal outcome types is additive.

- Additional Campaign Manager outcome signals beyond the four named in §6.3 are addable without breaking the existing signal contract; new signals MUST follow the same payload schema (including `ai_origin` and `emitted_at` fields).

13.6 Acceptance Criteria

The build of AI Concierge is complete when:

- [ ] Inbound Out-of-Hours and Overflow call handling works reliably for all configured modes.
- [ ] AI introduces itself as an AI on every call before substantive dialogue.
- [ ] Identity verification state machine (§3.3) is enforced; no patient-specific data is disclosed prior to `Verified`.
- [ ] Shared-number disambiguation triggers correctly where family linkage exists.
- [ ] Lead intake to Treatment Pipeline Manager occurs for new enquiries without duplication, with `SourceChannel`, `TriageStatus`, and `ConversationThreadID` populated and `BookingEligibility` initialised to `false`.
- [ ] BookingEligibility is read from Treatment Pipeline Manager and is never bypassed or inferred by AI Concierge.
- [ ] Rota-triggered short-notice recovery initiates outbound calls correctly and respects retry thresholds.
- [ ] Rota Manager availability is checked before each recovery workflow is initiated; negative tests (unavailable Rota Manager, no staff availability) result in Task Manager follow-up, not an outbound call.
- [ ] Suppression lifecycle signals (activation, lift, expiry) are emitted correctly to Campaign Manager per §6.1.
- [ ] All four Campaign Manager outcome signals are emitted with correct payloads (including `ai_origin` and `emitted_at`) at the correct lifecycle points; negative tests for batching and missing fields pass.
- [ ] Appointment confirmations and read-only record enquiries are supported post-verification (read-only; no write).
- [ ] All calls and recovery attempts generate auditable Call Threads with immutable transcripts.
- [ ] All Agent Console one-click actions are RBAC-enforced; negative tests (unauthorised access attempts) pass.
- [ ] AI boundaries in §7 are enforced; negative tests for prohibited AI actions pass.
- [ ] Audit log captures every event in §8 (including `RotaCancellationEvent` receipts and Rota Manager queries); logs are immutable and exportable.
- [ ] Task Manager follow-up is created for every unresolved call and every exhausted recovery workflow.
- [ ] Staff handover provides full context (transcript, intent summary, context card) and one-click actions.
- [ ] Telephony API traffic is routed through the External Provider API Gateway with Interactive and Background classes correctly segregated; a burst of Background-class recovery requests does not degrade Interactive inbound call handling.
- [ ] Performance Dashboard metric emissions include `ai_origin: true` and `emitted_at` on all events.
- [ ] All non-functional requirements in §14 are met.

14. Non-Functional Requirements

- **Performance:** Real-time transcription latency MUST be suitable for live staff takeover; transcript display in the Agent Console MUST not lag the live call by more than is perceptible during handover. Target AI

response latency for intent detection and first spoken response: to be defined by engineering based on telephony integration constraints.

- **Reliability:** Transcript continuity MUST be maintained during handover transitions. The module MUST degrade gracefully on telephony provider outage: if AI Concierge cannot answer, calls MUST be routed to staff queues or voicemail rather than dropped. Target availability: to be defined in the SLA.
- **Scalability:** The module MUST support multi-tenant deployment with tenant-isolated configuration surfaces (see §13.3). Recovery workflow volume spikes (e.g. mass rota cancellation) MUST not degrade inbound call handling performance; traffic class isolation via the External Provider API Gateway (§6.5) is the primary mechanism for achieving this.
- **Security:** Transcripts and call metadata MUST be encrypted at rest and in transit. RBAC enforcement via Access Manager applies to all data access and Agent Console actions. Secrets and API credentials for telephony integration (3CX and future providers) MUST be managed via the platform's secrets management facility, not stored in application configuration.
- **Privacy:** Transcript and call data are patient-bound personal data; GDPR rights (access, erasure, portability) apply. Data retention policy for Call Threads and transcripts MUST be configurable per tenant within platform-mandated bounds. AI Concierge MUST honour consent and contact preferences before initiating any outbound recovery call.
- **Observability:** The module MUST export: call volume metrics by mode and outcome; recovery workflow completion and exhaustion rates; identity verification success/failure rates; Agent Console action frequency by type. All metric emissions to Performance Dashboards MUST include `ai_origin: true` and an `emitted_at` freshness timestamp. Structured logs MUST include call_thread_id and contact_id for trace correlation. Distributed tracing MUST cover the inbound call path from telephony event to Call Thread creation.
- **Accessibility:** The Agent Console Calls Workspace MUST meet WCAG 2.1 AA standards. Spoken AI responses MUST use plain language and controlled vocabulary. Callers MUST be able to request human staff at any point in the dialogue.

15. Open Questions

1. **Short-notice window configuration defaults:** The spec states the default suppression expiry is 48 hours (tenant-configurable), but does not define the default for the short-notice recovery window itself (same-day only, or same-day + next-day?). The trigger definition references both as examples. This needs an authoritative default.
2. **Transcript retention policy:** The original is silent on how long Call Threads and transcripts are retained. The configurable retention surface (§13.3) is flagged, but the platform-mandated minimum and maximum bounds need definition before build.
3. **Telephony failover behaviour:** The spec states calls must not be dropped on AI Concierge outage, but the exact failover routing (staff queue vs. voicemail vs. IVR) is not defined. This needs an authoritative decision for the 3CX integration.
4. **Whitelisted patient record fields:** §4.1 references "whitelisted demographic or administrative fields" for read-only AI disclosure, but the whitelist content is not specified. The Practice Manager configuration surface defers this to runtime, but a safe default whitelist (or explicit statement that no fields are disclosed until explicitly whitelisted) is required before build.
5. **Transcript recording during staff takeover:** §4.5 and §13.3 both note this is configurable, but the default (on or off) is not stated in the original. Default-on vs. default-off has compliance implications and needs an authoritative decision.

6. **MFA requirements for sensitive operations:** §9 delegates MFA requirements to Access Manager policy. The specific operations (transcript export, configuration changes) that require MFA need to be formally declared in Access Manager's spec and cross-referenced here.

7. **AI response latency target:** §14 flags this as needing engineering definition based on telephony constraints. The target must be agreed between Product and Engineering before QA acceptance criteria can be finalised.

8. **Rota Manager availability API design:** §6.6 defines a sync read contract against Rota Manager for staff availability. The specific API endpoint, response schema, and SLA for this call need to be agreed with the Rota Manager module owner before the recovery eligibility check (§13.2, rule 18) can be implemented.

9. **`RotaCancellationEvent` schema ownership:** §6.2 defines the minimum required fields AI Concierge expects in the event payload. The canonical schema and versioning of `RotaCancellationEvent` are owned by Appointment Manager; any schema changes require an impact assessment against AI Concierge's inbound contract.