

Samay AI Agent - Requirements Document

1. Executive Summary

The **Samay AI Agent** (also referred to as "Samay AI Mentor") is a core differentiator of the Samay Rekha workforce management platform. It allows users (admins and employees) to interact with attendance data using natural language instead of manually navigating complex dashboards.

The agent must support queries in **English and Hindi**, provide **real-time insights**, and proactively **alert users to anomalies**.

2. Core Functional Requirements

2.1 Natural Language Query (NLQ) Engine

- **Goal:** Users can ask questions in plain language and receive accurate data-backed answers.
- **Supported Languages:** English, Hindi, Hinglish (mixed).
- **Key Use Cases:**
 - *Daily Status:* "Who is late today?", "Who is absent?", "Is Rajesh in the office?"
 - *Historical Analysis:* "How many days was Priya late last month?", "Show me the attendance trend for the Engineering team."
 - *Policy Queries:* "What is the policy for half-day leaves?", "How many sick leaves do I have left?"
 - *Anomaly Detection:* "Is anyone clocking out earlier than usual this week?" (Implied analysis).

2.2 Proactive Insights & Alerts

- **Goal:** The agent shouldn't just react; it should proactively notify admins of issues.
- **Triggers:**
 - Sudden spike in late arrivals.
 - Unusual absenteeism patterns.
 - "Ghost" employees (active status but no check-ins).
- **Delivery:** In-app notifications or "Morning Briefing" summary in the chat window.

2.3 Predictive Analytics

- **Goal:** Help admins plan better.
- **Capabilities:**
 - Forecast attendance for upcoming holidays/bridge days.
 - Suggest shift optimizations based on peak arrival times.

2.4 Behavioral Economics & Impact Analysis (The "Nudge" Engine)

- **Goal:** Change behavior through data visualization and "What If" scenarios.

- **Features:**
- **For Employees:**
 - *Salary Impact:* "If you continue coming 15 mins late every day, based on current policy, you might lose ₹20,000 this year."
 - *Career Impact:* "Consistently late employees are 30% less likely to get the top rating. Here is a research article on how punctuality affects trust."
 - *Long-term Projection:* "If this trend continues for 1 year, you will have missed 60 hours of work—that's 1.5 weeks of vacation time lost!"
- **For Admins:**
 - *Productivity Cost:* "Your team lost 45 hours this week due to late arrivals. That's equivalent to ₹50,000 in productivity."
 - *Management Suggestions:* "30% of your team is late on Mondays. Suggestion: Start Monday stand-ups at 10:30 AM instead of 9:30 AM to boost morale."
 - **Content Integration:** Link to genuine, high-credibility research (HBR, Forbes, SHRM) to back up claims.

2.5 Strategic HR Consulting & Global Context

- **Goal:** Act as a virtual HR Consultant, not just a data fetcher.
 - **Contextual Triggers:**
 - *Span of Control Check:* If total employees increase but manager count stays flat -> "Suggestion: You have 50 people reporting to 1 manager. Global best practices suggest a ratio of 8-10. Consider promoting team leads."
 - *Global Benchmarking:* "Your absenteeism rate is 5%. The industry average for Tech companies in India is 3%. You are doing better/worse than the community."
 - *News Feeds:* If a user asks about "Leave Policies", provide the answer based on internal policy BUT add: "By the way, 40% of top Indian startups are now offering 'Mental Health Days'. Here's a news snippet."
-

3. Technical Architecture for LLM Engineers

3.1 LLM Configuration

- **Model Class:** Low-latency, high-reasoning model (e.g., Gemini 1.5 Pro, GPT-4o).
- **Context Window:** Minimum 128k tokens to handle retrieval of weekly/monthly logs if necessary.
- **Temperature:** Low (0.2 - 0.3) for factual queries; Medium (0.5) for coaching/suggestions.

3.2 System Prompt & Persona (The "Desi Neighbour")

The agent should NOT sound like a robot. It should sound like a wise, trusted, and slightly informal "Neighbour" or "Elder Brother" (Bhaiya/Didi) who cares about your success.

- **Tone:** Warm, relatable, conversational, strictly "Desi" (Indian context).

- **Language Style:**
- *Code-Switching:* Use natural Hinglish (e.g., "Arre, look at the stats today!", "Matter serious hai.").
- *Idioms:* Use local metaphors (e.g., "Time is money, par peace of mind bhi zaroori hai.").
- *Empathy:* "Late ho gaye? Koi baat nahi, Mumbai traffic can be crazy. But let's try to be on time tomorrow for the meeting."
- **Safety:** Even while being informal, do not violate policy enforcement guidelines.
- **Constraint:** The "Neighbour" persona must be selectable. Corporate clients might prefer the "Professional" persona.

3.3 RAG (Retrieval-Augmented Generation) Strategy

The agent requires access to structured and unstructured data:

1. **Structured Data (SQL/NoSQL):**

- *Tables:* AttendanceLogs, Users, Offices, Leaves, Policies.
- *Access Method:* The LLM should NOT write SQL directly. It should call defined **Tools/Functions** (see Section 4).

1. **Unstructured Data (Vector DB):**

- *Content:* Company Handbooks, Leave Policies, Holiday Calendars.
- *Process:* Embed policy documents. On query, retrieve relevant chunks and feed to Context.

1. **External Knowledge Base (New):**

- *Sources:* Curated RSS feeds (HR News, Industry Stats), Psychology of Work papers.
 - *Usage:* When giving advice, retrieve 1 relevant external fact to validate the specific point.
-

4. Tool Definitions (Function Calling)

The LLM must be equipped with the following tools. Engineers must implement these API endpoints.

4.1 `get_daily_attendance`

- **Description:** Retrieves attendance status for a specific date or "today".
- **Parameters:**
- `date` (ISO8601 string)
- `filter_status` (enum: 'LATE', 'ABSENT', 'PRESENT', 'ON_LEAVE', 'ALL')
- `department` (optional string)

4.2 `get_employee_stats`

- **Description:** Gets detailed statistics for a specific employee over a period.
- **Parameters:**
- `employee_name` or `employee_id`
- `start_date`

- `end_date`

4.3 `get_policy_info`

- **Description:** Retrieves information from the company policy handbook.
- **Parameters:**
- `query` (string search term)

4.4 `compare_attendance_trends`

- **Description:** Compares attendance metrics between two time periods or teams.
- **Parameters:**
- `metric` (enum: 'LATE_ARRIVALS', 'ABSENTEEISM', 'OVERTIME')
- `team_a` (string)
- `team_b` (optional string)

4.5 `calculate_financial_impact` (New)

- **Description:** Calculates potential money lost due to late arrivals or absenteeism based on salary bands.
- **Parameters:**
- `entity_type` ('EMPLOYEE' or 'TEAM')
- `entity_id`
- `duration_months` (e.g., 12 for "1 year projection")

4.6 `get_external_hr_insight` (New)

- **Description:** Searches the external vector DB for research/news related to a topic.
 - **Parameters:**
 - `topic` (e.g., "benefits of punctuality", "employee-manager ratio")
 - `region` (default: "India")
-

5. UI/UX Specifications

5.1 Chat Interface

- **Floating Action Button (FAB):** Accessible from all screens (implemented).
- **Suggestions Bar:** "Chips" above the input field suggesting context-aware queries (e.g., "Who is late?", "My Leave Balance").
- **Thinking Indicator:** Visual feedback when the agent is querying the database.

5.2 Response Interpretation

- **Visual Data:** If the answer involves numbers, the agent should return a JSON payload that the UI renders as a **Chart** or **List** (as seen in the "Who came late today?" demo), rather than just text.
 - **Citations:** When answering policy questions, cite the source document.
-

6. Implementation Roadmap

Phase 1: The "Viewer" (Current Target)

- ☒ Basic Chat UI.
- ☒ Pre-defined queries mapped to hardcoded responses (Demo mode).
- ☐ Integration with `get_daily_attendance` API.
- ☐ Simple Rule-based NLP ("If 'late', call `get_late_users`").

Phase 2: The "Thinker" (LLM Integration)

- ☐ Connect to LLM Provider API.
- ☐ Implement RAG for User/Policy context.
- ☐ Deploy "Function Calling" layer for dynamic queries.

Phase 3: The "Predictor" (Advanced)

- ☐ Anomaly detection cron jobs running in background.
 - ☐ Push notification integration for alerts.
-

7. Security & Compliance

- **Data Masking:** PII (Personally Identifiable Information) must be handled securely. The LLM should process IDs, not raw names where possible, mapping them back at the UI layer.
- **Audit Logging:** Every query made to the AI must be logged for compliance auditing.
- **Hallucination Prevention:** Strict system prompts to prevent inventing employees or attendance events.