Product Spec – Agentic Customer Profile Capture (POC)

▼ Background and Context

Samar Capital will operate a RM led wealth management model, where long-term client relationships and contextual understanding will be central to advisory quality. Technology in this model is meant to augment RM effectiveness, not replace human judgment or force unnatural workflows.

A core input into advisory, compliance, and future intelligence will be the customer profile. In practice, this profile will not be built in a single step. Information about a client will emerge gradually across meetings, calls, messages, and informal conversations. Much of this information will be:

- Discovered in free-form discussion
- Captured as notes or voice dictation
- Not immediately mappable to a predefined field

Traditional CRM workflows struggle with this reality because they are optimized for structured, upfront data entry. As a result, valuable client context is often either lost or never recorded.

To address this, Samar Capital is exploring an agentic workflow for customer profile enrichment. The intent is to support progressive profiling, where an intelligent agent helps the RM:

- Ingest inputs from multiple sources (voice notes, text notes, meeting summaries, interview-style interactions)
- Extract structured information where possible
- Capture additional context as unstructured data or tags when it does not fit existing fields
- Identify gaps in the current customer profile and ask targeted follow-up questions at appropriate moments

The agent functions as a copilot that works alongside the RM, adapting to how information is naturally shared rather than forcing the RM into rigid forms.

This document defines a POC for this agentic workflow. The POC focuses on validating:

- Multi-modal input ingestion (voice, text, agent-led questions)
- Intelligent mapping of information to the current customer profile schema
- Safe capture and retention of additional unstructured client context
- Agent-driven prompting to progressively improve profile completeness over time

This POC is intentionally limited in scope and does not attempt to deliver full advisory logic, recommendations, or production-grade governance. Its primary goal is to validate the interaction model and data capture approach that can evolve alongside the customer profile schema over time.

▼ Problem Statement

Customer information at Samar Capital will not be collected in a single step. It will be discovered gradually through conversations, meetings, and follow-ups. Much of this information will be shared informally and captured as notes or voice dictation, if at all.

Typical CRM workflows are poorly suited to this reality. They rely on manual, form-based data entry and place the burden on RMs to decide what to record, where to record it, and when to update it. As a result, customer profiles are often incomplete, inconsistent, or outdated.

There is no system-level intelligence that:

- Helps extract usable information from unstructured RM inputs
- Retains important context that does not fit cleanly into structured fields
- Guides RMs on what customer information should be captured next

This leads to lost client context, higher downstream effort, and weaker foundations for advisory, compliance, and future intelligence.

This POC explores whether an agentic workflow can reduce these issues by assisting RMs in capturing customer information progressively, using a

combination of unstructured inputs and targeted follow-up questions.

▼ Goals & Success Criteria (POC)

▼ Goals

The goal of this POC is to validate an agentic customer profile capture system that can be built from scratch and supports how RMs naturally work.

Specifically, the POC should demonstrate that:

- An agent can ingest unstructured inputs (text and voice) generated by RMs
- The agent can extract relevant customer information from these inputs
- Structured data can be mapped to the current customer profile fields
- Additional information can be retained as unstructured notes or tags
- The agent can identify missing or unclear information and ask targeted follow-up questions
- Customer profiles can be incrementally enriched over multiple interactions

▼ What Success Looks Like

The POC is considered successful if:

- An RM can share notes or voice dictation without thinking about CRM fields
- The agent produces a structured output aligned to the customer profile
- The agent captures extra context without discarding it
- Follow-up questions are relevant, minimal, and non-intrusive
- Profile completeness improves over successive interactions

▼ Customer Profile Data Model (Current Schema + Extensible Context)

▼ Reference Schema

The current customer profile schema is defined in **Customer Profile – V2 (Google Sheets)**. This is the baseline set of structured attributes we want to

capture and progressively enrich over time.

▼ How the Excel Model Is Structured

Each row in the Excel represents **one CRM field**. The columns define how that field should be interpreted and captured:

- **Section**

  Logical grouping for readability and interview flow (e.g., Identity, Goals, Risk, etc.). Used to sequence the interview.

- **data_key**

  The canonical field identifier used in payloads and storage. This is the key engineering should treat as the "API name".

- **display_name**

  Human-friendly label for UI / agent prompts.

- **data_type**

  Storage and validation type. Observed types include:

  `text`, `date`, `enum`, `number`, `boolean`, `multi_select`, `json`.

- **priority**

  Capture urgency: `High / Medium / Low`. Used by the agent to decide what to ask now vs defer.

- **field_class**

  What kind of information the field represents (drives how strongly the agent should validate/challenge):

  - `FACT` (objective attributes, e.g., DOB)
  - `SIGNAL` (observations/behavioral indicators, e.g., preferences inferred)
  - `OPINION` (stated views, comfort levels, subjective inputs)
  - `SYSTEM` (system-generated / system-maintained fields)

- **source**

Expected provenance of the data (guides confidence and prompting style), e.g.:

`client_declared`, `rm_observed`, `questionnaire`, `system`, and combinations.

- **enum/options (incl. Other)**

  Allowed options for enum fields (including handling of `Other`).

- **notes / validation**

  Field-specific validation rules and capture guidance (e.g., date format, what to do when "Other" is selected, fallback rules like "use Age Band if DOB not available").

▼ Structured + Unstructured Profile Data

Not all useful customer information will map neatly into the current schema. The system must support **two parallel tracks** of profile capture:

1. **Structured profile fields**

   Data that cleanly maps to `data_key`s in the schema.

2. **Unstructured profile context**

   Any additional RM-provided context (notes, voice dictation, meeting learnings) that doesn't map to a field must still be retained as:

   - **Unstructured notes** (timestamped, source-tagged)
   - **Tags / topics** (lightweight labels extracted or added by RM/agent)

This ensures:

- no information is discarded,
- the schema can evolve later without losing historical context,
- future iterations can re-process old notes to backfill new fields.

▼ Model Evolution

This schema is the current representation of the customer profile and may evolve. The POC should therefore:

- map confidently when fields exist,
- store the rest as unstructured context,

▼ High-Level Workflow Overview

This POC implements an agentic customer profile enrichment workflow that operates across interactions, not as a one-time form-filling process. Each time new information about a customer is introduced, the same core loop is executed against the current state of the customer profile.

▼ Workflow Entry Points (Inputs)

The workflow can be triggered by any of the following RM-driven inputs:

- Text input (typed notes, pasted summaries)
- Voice input (voice dictation or meeting notes converted to text)
- Agent-led interview questions (triggered when clarification or missing data is identified)

All inputs are treated as unstructured and untrusted at the point of entry.

▼ Core Processing Loop

For every input, the system performs the following steps:

1. **Ingest**

   Normalize the input into text and attach metadata such as source, timestamp, RM, and customer identifier.

2. **Interpret**

   The agent analyzes the input to identify potential structured field values, additional contextual information, ambiguities, and conflicts.

3. **Compare**

   The extracted information is compared against the current customer profile to determine what is already known, what is missing, and what may require confirmation.

4. **Decide**

   Based on field priority and confidence, the agent decides whether to:

   a. Propose updates to structured profile fields

b. Store information as unstructured notes or tags

c. Enter interview mode to ask a clarifying or missing-information question

d. Defer capture to a future interaction

5. **Act**

   The agent either presents proposed updates for RM confirmation or asks a single, targeted interview-style question.

6. **Persist**

   Upon RM confirmation, data is saved as structured profile fields and/or unstructured customer context.

▼ Interview-Style Questioning

Interview mode is not a separate workflow. It is a conditional agent behavior invoked during the Decide and Act steps.

In interview mode:

- The agent asks one question at a time
- Questions are derived from missing or low-confidence customer profile fields
- Questions are contextual and conversational, based on prior inputs
- Each response is immediately processed through the same interpret, compare, and persist logic

Interview mode exits as soon as the required clarification is obtained or the RM chooses to defer the question.

> ⚠️ The POC does not assume a single fixed interview flow; interview-style questioning is expected to be adaptive and driven by missing or low-confidence profile data.

▼ Progressive Profiling Across Interactions

This workflow is stateful across time:

- Each interaction begins with the latest customer profile state
- New inputs incrementally enrich the profile
- Unstructured context is retained even when it does not map to current fields
- The agent's behavior adapts based on what is already known about the customer

No single interaction is expected to complete the customer profile.

▼ RM-in-the-Loop Control

At all stages:

- The RM can confirm, edit, or reject proposed updates
- Sensitive or ambiguous information is never auto-saved
- The RM remains the final authority on what is persisted

▼ Output of Each Interaction

Each interaction produces:

- Updated structured customer profile fields, if any
- Additional unstructured notes or tags
- A short summary of captured or proposed information
- An updated view of remaining gaps for future interactions

▼ Test and Seed Data (POC)

The POC will use a small set of synthetic or anonymized RM inputs to validate agent behavior. These inputs should represent different levels of completeness and ambiguity (e.g., partial notes, conflicting signals, missing high-priority fields). The specific test cases and data sets will be maintained separately and may evolve during implementation.

▼ Out of Scope (POC)

This POC does not include advisory logic, investment recommendations, risk or suitability calculations, or client-facing communication. It does not attempt to automate decision-making or finalize customer data without RM confirmation.

The scope is limited to validating agent-assisted capture and storage of customer information from RM inputs.

⚠️ **Footnote: Risk Profiling and Risk Bucketing**
While early discussions considered implementing risk profiling as a non-deterministic, agentic capability, this POC will not take that approach. Due to auditability and regulatory requirements, risk profiling and risk bucketing will be implemented using a deterministic model with clearly defined questions and scoring logic.
A separate specification (Risk Profiling - Form Based) defines the risk profiling questionnaire, scoring framework, and derivation of the final risk score or bucket. This POC may reference the outputs of that deterministic process but will not attempt to derive or infer risk profiles using non-deterministic agent logic.