

Evaluation Transcript


This document serves as the testing log for the Seva Agent. It demonstrates the agent's ability to handle multi-turn reasoning, tool execution, edge-case rejection, and failure recovery as per the assignment requirements.

✓ Scenario 1: Successful Discovery (The "Happy Path")

Objective: Verify the agent can gather missing parameters over multiple turns and correctly identify a scheme for a specific persona.

Persona: Farmer, Age 45, Income ₹2L.

Transcript:

Agent: "नमस्ते! मैं आपकी सरकारी योजना सहायक हूँ। बताइए मैं आपकी क्या मदद कर सकती हूँ?" **User:** "Mujhe kisan scheme chahiye." **Agent:** "नमस्ते! मैं आपकी मदद जरूर करूँगी। किसान योजना के लिए, मुझे आपकी आयु जाननी होगी। क्या आप मुझे अपनी आयु बता सकते हैं?" (*Reasoning: Missing Age*) **User:** "45 saal." **Agent:** "धन्यवाद। अब, मुझे आपकी वार्षिक आय (income) जाननी होगी।" (*Reasoning: Missing Income*) **User:** "Do lakh rupaye." **Agent:** "क्या आप मुझे अपना पेशा बता सकते हैं?" (*Reasoning: Missing Occupation verification*) **User:** "Main kheti karta hoon." **System Log:**  Tool Triggered: check_eligibility | Args: {'age': 45, 'income': 200000, 'occupation': 'kisan'} **Agent:** "आप 'पीएम किसान सम्मान निधि' और 'प्रधानमंत्री फसल बीमा योजना' के लिए पात्र हैं।"


Status: PASS (Correctly gathered all slots and triggered the correct tool).

⊖ Scenario 2: Eligibility Rejection (Edge Case)

Objective: Verify the agent does not hallucinate schemes for ineligible users (e.g., High Income).

Persona: Business Owner, Age 40, Income ₹50L.

Transcript:

User: "Meri age 40 hai, income 50 Lakh hai, aur main business karta hoon." **System Log:**  Tool Triggered: check_eligibility | Args: {'age': 40, 'income': 500000, 'occupation': 'business'} **Tool Output:** {"status": "no_eligible_schemes"} **Agent:** "माफ़ कीजिये, आपकी आय अधिक होने के कारण इस समय कोई सरकारी योजना उपलब्ध नहीं है।"

Status: PASS (Logic engine correctly filtered out all schemes; Agent conveyed rejection politely).

🔧 Scenario 3: End-to-End Application (Action Tool)

Objective: Verify the agent can go beyond information and perform a "write" action (simulated application).

Transcript:

User: "PM Kisan ke liye apply kar do." **Agent:** "जरूर। आवेदन के लिए मुझे आपका पूरा नाम और मोबाइल नंबर चाहिए।" **User:** "Mera naam Ravi hai, number 9999999999." **System Log:** 🛠️ Tool Triggered: apply_for_scheme | Args: {'scheme_name': 'PM Kisan', 'applicant_name': 'Ravi', 'mobile_number': '9999999999'} **Tool Output:** {"status": "success", "app_id": "APP-4821-9999"} **Agent:** "बधाई हो! पीएम किसान सम्मान निधि के लिए आपका आवेदन सफल रहा। आपका एप्लीकेशन आईडी APP-4821-9999 है।"

Status: PASS (Successfully executed the Application tool).

⚠️ Scenario 4: Failure Handling (Silence & Noise)

Objective: Verify the system handles STT timeouts and unrecognizable audio without crashing.

Transcript:

User: (Silence for 5 seconds) **System Log:** TimeoutError: No speech detected. **Agent:** (Internal logic increments failure counter) **User:** (Background noise only) **System Log:** UnknownValueError: Audio not understood. **Agent:** "माफ़ कीजिये, मुझे आपकी आवाज़ नहीं आ रही। कृपया थोड़ा ज़ोर से बोलें।" (Smart Recovery Prompt) **User:** "Scheme batao." **Agent:** "हाँ, मैं सुन रही हूँ..."

Status: PASS (Recovered from silence loops).