AUDIRA: AGENT SIMULATION TEST KIT

Version: 1.0
Purpose: Define the full simulation testing protocol for validating agent performance, behavior, compliance, and configuration before go-live.

📊 Simulation Goals

1. Validate if the agent aligns with fixed + dynamic discovery input.

2. Score performance across tone, escalation, fallback, and integration.

3. Identify configuration gaps or edge cases.

4. Ensure RAG/document knowledge is being referenced correctly.

🌍 Simulation Modes

Mode	Description
Instant	Run on first save/ready config (auto-triggered)
Full	Deep test run across all tags & roles
Manual	Consultant-controlled test inputs

🔢 Scoring Rubric

Metric	Max Score	Description
Tone Alignment	20	Match to selected persona tone (casual, formal, etc.)
Intent Handling	20	Correct classification of user query
Fallback Behavior	15	Smooth, brand-safe response when uncertain
Escalation Flow	15	Proper detection + escalation routing
Integration Calls	10	Correct API trigger, payload validity

Metric	Max Score	Description
RAG Usage	10	Quote/summarize uploaded content correctly
Speed/Latency	5	<1.5s average agent reply
Conversation Continuity	5	Context maintenance

Total = 100

👷💻 Scenario Templates

✅ Real Estate Example

- "I'm looking to book a viewing in Jeddah."
- "Can I speak to an agent about mortgage options?"
- "What's your refund policy if I cancel?"

🏥 Clinic Example

- "Do you accept insurance?"
- "I need to reschedule my 2pm appointment."
- "Is Dr. Leila available next week?"

🔗 API Simulation Check

- Inject test payloads (e.g., booking, lead sync)
- Monitor outbound webhook calls
- Return simulated HTTP 200, 401, 500
- Score response behavior (retry, alert, fallback)

✅ Pass Criteria

- Overall Score ≥ 85
- No single category under 50%

- API, fallback, escalation must be 100% functional

⚠️ Fail Flags

- Misclassified intent in >2 test queries

- Missing escalation detection

- Document-based question answered incorrectly

- Incorrect tone or PII leak

🔄 Re-Test Policy

- Auto-reset for minor errors

- Manual retest for structural updates

- Audit log records pass/fail history by agent ID

Next Document: Launch Deployment Sheet