FF FUT

SAMPLE ·

The Sev

Completed by:

ver the Horizon

	GOAL CONTEXT What outcome is the agent pursuing?	✓ *Goal: Classify FNOL claim type, assign adjuster queue, draft ack letter within 4 min.*			
		Does the agent know when it has succeeded (a measurable completion criterion)? ✓ *Success: triage accuracy ≥92%, cycle time <4 min, zero misrouted high-severity claims.*	YES		
		Is the goal reviewed and updated when business priorities change? ~ *Reviewed quarterly. Last update Feb 2026. Formal change process not yet documented.*		PART	
2	**2** **IDENTITY CONTEXT** Who is the agent acting as?	Is the agent's persona and authority level explicitly defined? ✓ *Agent acts as 'Claims Intake Coordinator' — read/classify/route only. No settlement authority.*	YES		
		Does the agent know which team or function it represents? ✓ *Represents Claims Operations. Cannot represent Legal, Finance, or Executive functions.*	YES		
		Is there a human owner the agent can be attributed to in any output? ✓ *All outputs attributed to 'AcmeCorp Claims Operations — AI-Assisted' with human review flag.*	YES		
3	**3** **PERMISSION CONTEXT** What is the agent allowed to read, write, spend, or send?	Are read permissions scoped to only the data the agent needs? ✓ *Agent reads: claim form, policy record, claimant history. No access to financial reserves.*	YES		
		Are write/send/spend actions bounded by explicit limits (e.g., max $ per transaction)? ✓ *Agent writes to routing queue only. No financial transactions. Ack letter requires human approval.*	YES		
		Is there a deny-list of actions the agent must never take? ✓ *Deny-list: no settlement offers, no coverage denials, no communication with legal counsel.*	YES		
4	**4** **SYSTEM CONTEXT** What tools, APIs, and data stores are available?	Is the full tool inventory documented and versioned? ✓ *Tools: Guidewire ClaimCenter API v2.1, internal policy DB, routing queue API. All documented.*	YES		
		Does the agent gracefully handle tool unavailability without hallucinating a result? ~ *Guidewire fallback tested. Policy DB fallback not yet tested. Added to pre-launch checklist.*		PART	
		Are tool credentials rotated and access logged? ✓ *API keys rotated every 90 days. All access logged to SIEM with agent identity tag.*	YES		
5	**5** **HISTORY CONTEXT** What has happened in this case / customer / workflow before?	Does the agent have access to relevant prior interactions for this entity? ✓ *Agent reads last 24 months of claim history for the policy holder before triage.*	YES		
		Is history bounded (e.g., last 90 days) to avoid context-window overload? ✓ *History window: 24 months claims, 6 months interactions. Tested at p95 context size.*	YES		
		Is historical data quality-checked before being passed to the agent? ~ *Quality check on completeness (>97%) but not on consistency across source systems.*		PART	
6	**6** **POLICY CONTEXT** What rules, regulations, and exception paths apply?	Are applicable regulations (GDPR, HIPAA, SOX, etc.) encoded as explicit constraints? ✓ *State insurance regulations (CA, TX, NY, FL) encoded. CCPA data minimization enforced.*	YES		
		Does the agent know the exception path when a rule cannot be automatically satisfied? ✓ *Exception path: escalate to Senior Adjuster queue. SLA: human review within 2 hours.*	YES		
		Is the policy layer versioned and auditable independently of the model? ~ *Policy rules in config file (versioned in Git). Audit trail exists but not automated.*		PART	
7	**7** **ESCALATION CONTEXT** When and to whom does the agent hand off?	Is there a defined confidence threshold below which the agent must escalate? ✓ *Escalate if classification confidence <85% OR claim value >$50K OR injury claim detected.*	YES		
		Is the escalation path tested — does a human actually receive and act on escalations? ✓ *Escalation tested in staging. Senior Adjuster queue monitored. SLA: 2-hour response.*	YES		
		Is escalation latency measured and within SLA? ~ *Latency measured in staging (avg 47 min). Production SLA monitoring not yet live.*		PART	
	SCORE BOX AcmeCorp result: 16 Yes · 5 Partial · 0 No				