

SAMPLE · FILLED EXAMPLE · ACMECORP INSURANCE

AcmeCorp Claims-Triage Agent · Pilot Review: May 2026

Pilot-to-Production Graduation Rubric — Sample

The Six Graduation Criteria — AcmeCorp Assessment

Reviewer: David Park, CRO · Agent: Claims-Triage v1.2 · Pilot period: Feb 15 – May 9, 2026 · Claims processed: 4,847

	CRITERION	DEFINITION	EVIDENCE	OWNER	STATUS	ASSESSOR NOTES
1	Outcome Reliability	Agent meets or exceeds target accuracy on a held-out test set that was NOT used during development.	Eval report: 94.2% triage accuracy on 500-claim held-out set (target: ≥92%). Signed: Sarah Okonkwo, ML Lead, May 9.	ML/Data Lead	GREEN	Exceeds 92% target. Misclassification analysis shows remaining 5.8% are ambiguous multi-peril claims — acceptable.
2	Cost Per Task	Fully-loaded cost per agent task (inference + infra + human review) is within the approved budget envelope.	Cost model: \$1.84/claim fully loaded (inference \$0.42 + infra \$0.31 + 8% human review overhead). Budget: \$2.50. Signed: CFO.	Finance / Ops	GREEN	\$1.84 vs \$2.50 budget. 26% under budget. Annualized savings vs manual: \$487K at current volume.
3	Latency at P95	95th-percentile response time is within the agreed SLA for the business process.	Load test (1,000 concurrent): p95 = 3.2 min. SLA: <4 min. Tested at 1.5× peak load. Signed: James Liu, Eng Lead.	Engineering	GREEN	p95 3.2 min well within 4-min SLA. p99 = 3.8 min (still within SLA). Headroom confirmed at 1.5× load.
4	Audit Trail Completeness	Every agent action is logged with: timestamp, input, output, model version, and tool calls made.	Audit coverage report: 99.97% of actions logged (3 gaps in 4,847 claims — all during Feb 18 maintenance window).	Security / Compliance	YELLOW	99.97% coverage. 3 gaps traced to maintenance window — not a systemic issue. Monitoring alert added. Conditional GREEN.
5	Off-Ramp Tested	The procedure for reverting to the human process has been executed in a staging environment and documented.	Off-ramp runbook v1.1. Staging test March 14: manual queue activated in 4 min 12 sec. Signed: Ops Lead.	Ops Lead	GREEN	Off-ramp tested twice in staging. Activation time: 4:12 (target <5 min). Runbook published to ops wiki.
6	Human-in-the-Loop Policy Ratified	The policy defining which actions require human approval has been reviewed and signed by Ops and Legal.	HITL Policy v1.0 signed by Maria Chen (VP Ops) and Rachel Torres (General Counsel) on April 28, 2026.	Ops + Legal	GREEN	Policy covers: injury claims, >\$50K, fraud flags, <85% confidence. Exception log template deployed to Jira.

Go / No-Go Decision — AcmeCorp Claims-Triage Agent

5 GREEN, 1 YELLOW (conditional) → CONDITIONAL GO. Condition must be resolved within 30 days of launch.

CRITERION	STATUS	CONDITION / FIX REQUIRED
Outcome Reliability	GREEN	None — exceeds target.
Cost Per Task	GREEN	None — 26% under budget.
Latency at P95	GREEN	None — headroom confirmed at 1.5× load.
Audit Trail Completeness	YELLOW	Add monitoring alert for gaps >0 per 24-hr window. Due: June 1, 2026. Owner: James Liu.
Off-Ramp Tested	GREEN	None — runbook published and tested.
HITL Policy Ratified	GREEN	None — signed by Ops and Legal.

OVERALL DECISION: CONDITIONAL GO ✓

5 of 6 criteria GREEN. Criterion 4 (Audit Trail) is YELLOW — monitoring alert must be live within 30 days.
Launch authorized for: May 19, 2026 · Launch owner: Maria Chen, VP Operations
Condition resolution deadline: June 1, 2026 · Condition owner: James Liu, Engineering Lead

SIGNATURES

David Park
Chief Risk Officer
May 12, 2026

Maria Chen
VP Operations
May 12, 2026

Rachel Torres
General Counsel
May 12, 2026

James Liu
Engineering Lead
May 12, 2026