

DECISIONFLOW

# AI Operations and Model Governance Report

*Production evaluation, model and prompt change control, routing, provenance, cost, drift, incident, and rollback*

DELIVER PHASE | Version 1.0 | Illustrative 90-day controlled-scale program

EVIDENCE & PUBLISHING INTEGRITY

This public portfolio edition distinguishes actual evidence, assumed continuity steps, simulated case-study outcomes, targets, and recommendations. Simulated customers, commercial results, benchmarks, incidents, quotations, and later-lifecycle outcomes are illustrative and must not be interpreted as results from a live commercial product. Evidence labels are retained throughout the package to support transparent and responsible interpretation.

Document Control

| Control              | Definition                                                                                                                             |
|----------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Document code        | DEL-04                                                                                                                                 |
| Version / status     | 2.0 — Final Public Portfolio Edition — Simulated Delivery Scenario                                                                     |
| Baseline / author    | 14 July 2026 - Min Thu Kyaw                                                                                                            |
| Lifecycle continuity | Actual MVP -> DEV-06 assumed productionization -> simulated launch -> Conditional Scale -> Deliver phase.                              |
| Delivery decision    | Execute a 90-day controlled-scale program; keep collaboration, integrations, richer ingestion, and enterprise features evidence-gated. |
| Evidence rule        | Actual, assumed, simulated, target, and recommendation statements are explicitly separated.                                            |

Actual Project Artifacts

| Artifact                | Reference                       |
|-------------------------|---------------------------------|
| Low-Fidelity Prototype  | <a href="#">View prototype</a>  |
| High-Fidelity Prototype | <a href="#">View prototype</a>  |
| MVP Frontend Repository | <a href="#">Open repository</a> |
| MVP Backend Repository  | <a href="#">Open repository</a> |

# Contents

|    |                                                         |
|----|---------------------------------------------------------|
| 01 | <a href="#">AI operating objective and boundaries</a>   |
| 02 | <a href="#">Production AI workflow</a>                  |
| 03 | <a href="#">Evaluation and promotion gates</a>          |
| 04 | <a href="#">Model routing and cost governance</a>       |
| 05 | <a href="#">Monitoring and drift</a>                    |
| 06 | <a href="#">Human oversight and user authority</a>      |
| 07 | <a href="#">Incident and rollback process</a>           |
| 08 | <a href="#">Delivery result and residual conditions</a> |

# AI Operating Objective and Boundaries

DecisionFlow AI operations optimize for grounded, reviewable decision support rather than autonomous decision-making. Quality, provenance, cost, latency, and failure behavior are managed as production product responsibilities. Model agreement with a preferred answer is not a quality metric.

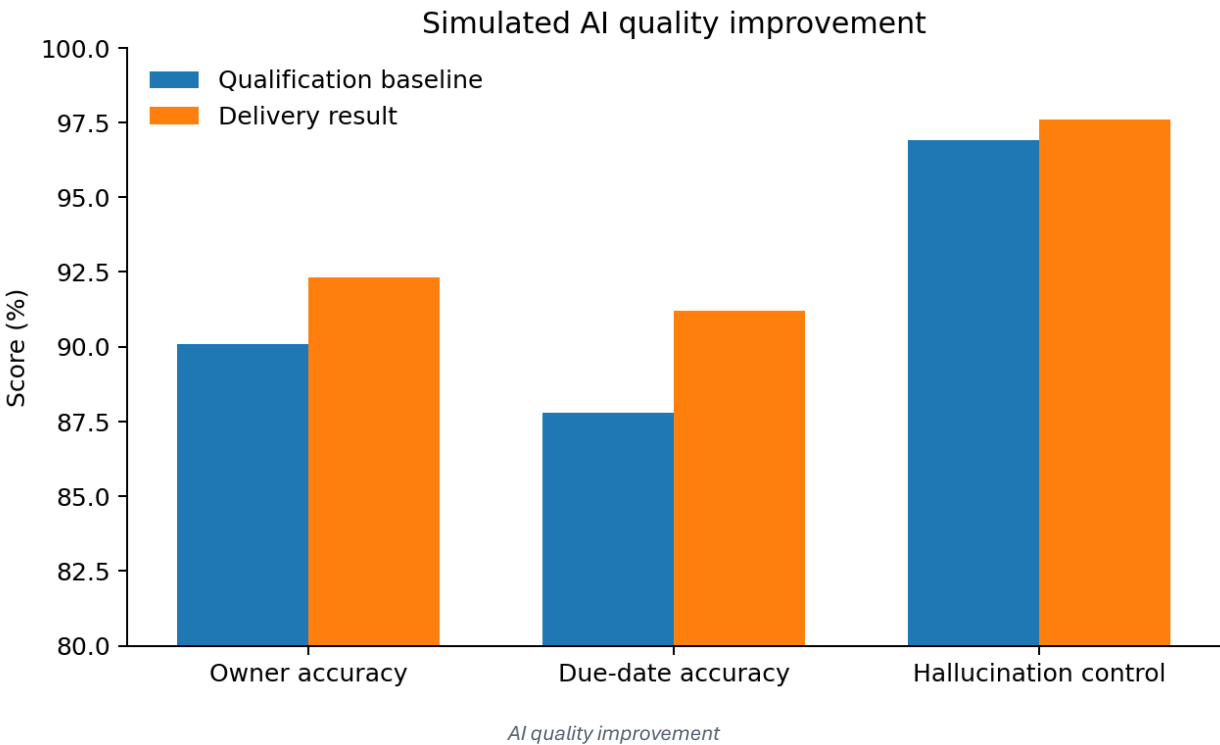
**Non-negotiable boundary** - AI may structure, compare, recommend, explain, and identify uncertainty. The authenticated user must review and explicitly finalize the decision. No model or routing policy may remove this authority.

## Production AI Workflow

- 1. Validate content type, size, encoding, control characters, ownership, and consent notice.
- 2. Create an idempotent extraction job with source, decision ID, prompt version, routing policy, and requested output schema.
- 3. Call the selected provider/model, parse JSON, validate schema, and store provenance, usage, latency, and fallback state.
- 4. Present editable extraction and require confirmation for accountability fields.
- 5. Generate a versioned recommendation from reviewed values, criteria, notes, and saved weights.
- 6. Show evidence, confidence, risks, alternatives, provenance, and any fallback label.
- 7. Record user edits, regeneration, acceptance, override, final option, and finalization event without treating agreement as quality proof.

| AI artifact    | Versioned fields                                                                                                                      |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Extraction     | Provider, model, prompt version, schema version, routing policy, source hash/reference, usage, latency, fallback, evaluation cohort.  |
| Recommendation | Reviewed input version, criteria/weights version, provider, model, prompt, confidence, risks, alternatives, usage, latency, fallback. |
| Final decision | Selected option, accepted recommendation version if any, user edits, actor, timestamp, immutable snapshot.                            |

## Evaluation and Promotion Gates

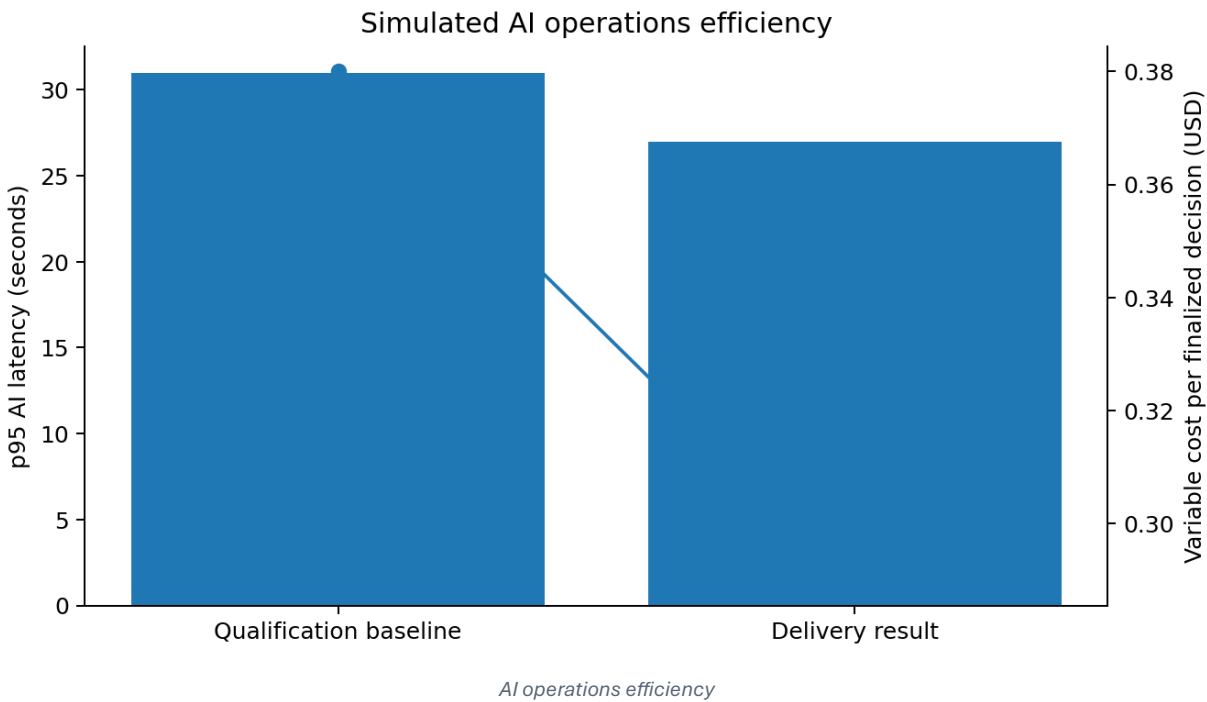


| Evaluation dimension | Qualification baseline | Promotion gate | Simulated delivery result | Decision |
|----------------------|------------------------|----------------|---------------------------|----------|
|----------------------|------------------------|----------------|---------------------------|----------|

| Evaluation dimension               | Qualification baseline | Promotion gate        | Simulated delivery result | Decision |
|------------------------------------|------------------------|-----------------------|---------------------------|----------|
| Owner accuracy                     | 90.1%                  | >=90%                 | 92.3%                     | Pass     |
| Due-date accuracy                  | 87.8%                  | >=90% or reduce claim | 91.2%                     | Pass     |
| Hallucination control              | 96.9%                  | >=95%                 | 97.6%                     | Pass     |
| Confidence calibration ECE         | 0.12                   | <=0.10                | 0.08                      | Pass     |
| Fallback rate                      | 1.5%                   | <=2%                  | 0.9%                      | Pass     |
| Variable cost / finalized decision | \$0.38                 | <=\$0.32              | \$0.29                    | Pass     |
| p95 AI latency                     | 31s                    | <=35s at normal load  | 27s                       | Pass     |

- The fixed evaluation set includes complete transcripts, sparse notes, conflicting criteria, prompt injection, sensitive-data cues, no-option cases, relative dates, ambiguous speakers, and supported English-domain samples.
- Every prompt or model candidate is compared against the production version using the same fixtures and scoring rubric.
- A model change fails promotion if any critical safety case fails, if a key metric crosses its floor, or if cost/latency improvement depends on hidden quality loss.

Model Routing and Cost Governance



| Routing condition                        | Preferred route                                                                          | Guardrail                                                |
|------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------|
| Short, well-structured English source    | Cost-efficient validated model                                                           | Same extraction schema and minimum field-quality floors. |
| Long, ambiguous, or high-conflict source | Higher-capability validated model                                                        | Content-size and budget controls; visible progress.      |
| Provider health degraded                 | Approved alternate or actionable failure                                                 | Never silently change provenance or quality policy.      |
| No safe/qualified route                  | Do not generate; explain next action                                                     | No unlabeled heuristic recommendation.                   |
| Regeneration                             | Same or escalated route based on failure reason                                          | User sees version and cost/latency expectation.          |
| Cost control                             | Implementation                                                                           |                                                          |
| Budget                                   | Per-user and environment caps with alerts and explicit handling when limits are reached. |                                                          |
| Input efficiency                         | Character/token limits, duplicate-submission prevention, and reviewed-input reuse.       |                                                          |
| Output efficiency                        | Constrained schemas and concise reasoning fields rather than unrestricted prose.         |                                                          |
| Routing                                  | Quality-gated model selection by complexity, not lowest cost alone.                      |                                                          |

| Routing condition | Preferred route                                                                         | Guardrail |
|-------------------|-----------------------------------------------------------------------------------------|-----------|
| Monitoring        | Cost per generation, regeneration, finalized decision, active user, and paying account. |           |

## Monitoring and Drift

| Signal                     | Warning / action threshold     | Action                                                                                    |
|----------------------------|--------------------------------|-------------------------------------------------------------------------------------------|
| Schema-validation failures | >1% over qualified sample      | Freeze promotion; inspect provider/model change and prompt.                               |
| Fallback rate              | >2%                            | Investigate provider health and parser failures; communicate degraded state.              |
| Unsupported material claim | >5% or any high-severity case  | Freeze affected model/prompt and review outputs.                                          |
| Owner or due-date accuracy | <90%                           | Mandatory confirmation remains; retrain prompt/evaluation and reduce claim if persistent. |
| Calibration ECE            | >0.10                          | Recalibrate score/copy and restrict confidence display.                                   |
| Variable cost              | >\$0.32 per finalized decision | Review routing, regeneration and output size.                                             |
| p95 latency                | >35s normal or >45s at 40 jobs | Review queue, provider and capacity; reduce growth cap.                                   |

## Human Oversight and User Authority

| Control                     | Required behavior                                                                                                |
|-----------------------------|------------------------------------------------------------------------------------------------------------------|
| Editable AI output          | Topic, options, criteria, stakeholders, risks, actions, owners and dates remain reviewable.                      |
| Accountability confirmation | Owner and due date require explicit confirmation or correction before finalization.                              |
| Recommendation transparency | Evidence, reasoning, confidence, risks, alternatives, model/prompt provenance and fallback state are visible.    |
| Override support            | User may select a different final option and record rationale without penalty.                                   |
| Immutable final record      | Final snapshot preserves AI version, user edits, selected option, actor and time.                                |
| High-stakes limitation      | Product copy states that DecisionFlow is advisory and not a substitute for professional or accountable judgment. |

## Incident and Rollback Process

8. Detect through evaluation, monitoring, support, or operator review and classify severity.
9. Freeze model/prompt/routing promotion and preserve affected version and evidence.
10. Determine affected users, decisions, time window, provider, model, prompt and failure mode.
11. Rollback to the last qualified version or disable generation when no safe version exists.
12. Communicate actionable status without exposing other customer content.
13. Correct affected records when appropriate, document the decision, and add a regression fixture.
14. Resume only after evaluation and release gates pass.

## Delivery Result and Residual Conditions

**AI operations decision** - QUALIFIED FOR CONTINUED CONTROLLED SCALE in English-language, human-reviewed individual workflows. Quality, calibration, cost, normal latency, and the 40-job gate pass in the simulated program.

- The 80-job profile remains outside the approved operating boundary: 68s p95 and 1.4% failures.
- One primary provider remains a material dependency even with an alternate route and game-day procedure.
- Commercial multilingual quality claims remain unsupported.
- High-stakes decisions continue to require explicit human responsibility and domain-appropriate review.

*Evidence discipline: replace all simulated delivery values with observed, reproducible results before presenting them as actual product performance.*