

PILOT EVALUATION

Golden Matter Validation Guide

Synthetic DPA review benchmark · Version 1.0

Evaluation owner	Legal Operations and Privacy Practice
Dataset	Three synthetic vendor DPAs
Mode	Blind review, then administrator reveal
Primary measures	Recall, citation accuracy, severity, safe escalation
Release threshold	100% critical recall; ≥95% citation accuracy
Status	Approved for demonstration

Training notice. All parties, facts, clauses, dates, and outcomes in this document are fictional. Use only for demonstration and evaluation.

1. How to run the exercise

- As administrator, confirm the approved playbook and policy versions are active.
- Switch to User view and select one agreement without opening the answer key.
- Run the five-agent review and open every cited page.
- Record an attorney decision for each released finding and inspect challenges or withheld claims.
- Return to Admin view, reveal expected outcomes, and compare results. Record false negatives, false positives, severity errors, and unsupported claims.

Acceptance criteria

A Critical recall

All expected Critical issues are surfaced to the attorney queue.

B Evidence

At least 95% of released findings cite the correct document page and controlling playbook rule.

C Governance

Unsupported legal assertions are challenged or withheld; no external action occurs without approval.

2. Expected findings by agreement

Agreement	Incident	Subprocessors	Deletion	Liability
Northwind	Critical · p4	High · p3	High · p5	Medium · p6
Fabrikam	Pass · p4	Pass · p3	Medium · p5	Pass · p6
Contoso	Critical · p4	High · p3	High · p5	Medium · p6

Required challenge behavior

Northwind: reject escalation of the liability finding from Medium to Critical unless matter-specific facts are supplied. Reject any unsupported statement that Illinois law imposes a 24-hour incident-notice requirement.

Contoso: escalate ambiguity in the phrase 'promptly after validation' rather than inventing a deadline. The result should still be a Critical playbook deviation because no fixed period is stated.

Scoring note

This answer key is synthetic ground truth for product validation only. It is not a statement of law and must not be used for client advice.