

01 | BOUND

Bound the decision and the authority

Choose one consequential workflow. Define what AI may contribute, what must be evaluated, and who is actually authorized to act before discussing broader automation.

ORGANIZATION / TEAM

WORKFLOW OWNER

ASSESSMENT DATE

VERSION

01 Write the authority contract

AI may propose

CONTRIBUTION

for the decision to

CONSEQUENTIAL ACTION

but only

NAMED ROLE OR EXTERNAL PROCESS

may authorize

EXACT ACTION OR COMMITMENT

02 Why this decision is consequential

☐ Affects rights, access, or treatment

☐ Commits money, capacity, or terms

☐ Changes safety or operating posture

☐ Creates a regulated or public action

☐ Is difficult to reverse or appeal

☐ Can propagate across systems

ASSESSMENT BOUNDARY

This aid exposes responsibilities and frames a bounded evaluation. It does not establish legal authority, determine compliance, validate a policy, certify an AI system, or prove that a proposed control is effective.

03 Assign the responsibilities

Responsibility	May do	Must not do	Named owner	Confirmed?
AI assistant or agent				Yes / No / ?
Governed model				Yes / No / ?
Human reviewer				Yes / No / ?
External authority / process				Yes / No / ?

04 Draw the absolute boundary

THE SYSTEM MUST STOP BEFORE

EXISTING FALLBACK / APPEAL / ESCALATION

ACTION THAT ONLY THE ACCOUNTABLE AUTHORITY MAY TAKE

THE FIRST USEFUL EVALUATION OUTCOME WOULD BE

Keep unlike roles distinct: AI proposes - the model evaluates - a person authorizes - the record preserves what happened.

Trace one proposal through the system

Retain the AI contribution with attribution. Then connect it to approved facts, explicit model logic, an accountable decision, and a record another reviewer can reconstruct.

01 Identify the AI contribution

SYSTEM / MODEL

PROMPT / INSTRUCTION REVISION

CONTEXT REVISION

PROPOSAL REVISION

WHAT AI SUPPLIED

KNOWN LIMITS OR UNRESOLVED CLAIMS

02 Resolve the important facts

Fact used by proposal	Approved source	Revision	As-of time	State	Owner
				Confirmed / stale / missing	
				Confirmed / stale / missing	
				Confirmed / stale / missing	

03 Name the evaluation model

MODEL / POLICY REVISION

APPROVED OWNER

VALIDATION / ACCEPTANCE CASES

Rule, calculation, or constraint	Source / authority	Revision	Reason it applies	Tested?
				Yes / No / ?
				Yes / No / ?

04 Follow the responsibility sequence

01AI proposal

02Approved facts

03Model evaluation

04Human decision

05Record / outcome

KEEP UNLIKE CLAIMS SEPARATE

Source fact

Assumption

AI proposal

Modeled finding

Recommendation

Authorized decision

Observed outcome

05 State what the model found - and what it cannot conclude

MODELED FINDING AND REASON

MISSING INFORMATION / UNCERTAINTY

WHAT REMAINS FOR THE ACCOUNTABLE PERSON TO JUDGE

03 | REVIEW

Design the review packet and safeguards

Human authority is meaningful only when the reviewer can inspect the basis, disagree with the proposal, use an escalation path, and act within a recognized scope.

01 Define what the reviewer must see

Review element	How it will be shown	Evidence source	Who verifies it	Required?
Proposal and AI contribution				Yes / No
Facts, sources, and freshness				Yes / No
Rules, constraints, and model revision				Yes / No
Alternatives, reasons, and uncertainty				Yes / No
Authority, exception, and appeal path				Yes / No

02 Make human review real

REVIEWER MUST BE ABLE TO

☐ Approve within a named scope

☐ Narrow the proposed action

☐ Request or correct evidence

☐ Choose another represented option

☐ Defer until conditions are met

☐ Reject and record a reason

REVIEW CONDITIONS ARE REAL

☐ Enough time for the consequence

☐ Relevant domain competence

☐ Permission to disagree

☐ Accessible evidence and interface

☐ No misleading default action

☐ Coverage for escalation

03 Confirm safeguards outside the AI session

☐ Identity and permission controls

☐ Security and privacy review

☐ Fallback or rollback path

☐ Monitoring for drift / missing data

☐ Source and rule ownership

☐ Bias review where relevant

☐ Appeal or incident response

☐ External action separately authorized

☐ Approved correctness cases

☐ Override and exception record

☐ Change review for model / prompts

☐ Outcome separate from decision

04 Name the stop, fallback, and escalation paths

ESCALATION TRIGGER

ESCALATION OWNER / CHANNEL

FALLBACK WHEN AI OR MODEL IS UNAVAILABLE

PROHIBITED AUTOMATED ACTION

CONDITION THAT REQUIRES A STOP

EXISTING LEGAL / POLICY / SAFETY PROCESS

DO NOT TREAT PROCESS DESIGN AS OUTCOME EVIDENCE

Attribution, validation, explanation, and authorization can make a decision inspectable. They do not prove the source complete, the policy appropriate, the decision lawful or fair, or the eventual outcome good.

04 | EVALUATE

Run a bounded evaluation and decide readiness

Test one useful slice under controlled conditions. Record the baseline before choosing targets, keep external authority in place, and decide whether to proceed, narrow, prepare, or stop.

01 Write the evaluation statement

NAMED DECISION / USERS

allow AI to

BOUNDED CONTRIBUTION

For

evaluate with

APPROVED FACTS AND MODEL

while retaining

EXISTING AUTHORITY / FALLBACK

MODE: HISTORICAL / SYNTHETIC / SHADOW / LIMITED LIVE

INCLUDED CASES / PERIOD

GO / NO-GO AUTHORITY

02 Choose measures before observing results

Measure	Current baseline	Observation method	Desired criterion	Observed result
Correctness against approved cases				
Unsupported or stale facts detected				
Reviewer can reconstruct the finding				
Human correction / deferral / rejection				
Workflow safety or quality measure				

CASES THAT MUST PASS

ADVERSARIAL, MISSING-DATA, AND EDGE CASES

03 Score readiness

Dimension	Ready	Needs work	Not yet	Evidence or next action
Decision and authority boundary	[]	[]	[]	
AI contribution and attribution	[]	[]	[]	
Approved facts, rules, and cases	[]	[]	[]	
Reviewer capability and safeguards	[]	[]	[]	
Evidence record and outcome observation	[]	[]	[]	

DISPOSITION: PROCEED / NARROW / PREPARE / DISCOVER / NOT APPROPRIATE

NEXT ACTION / OWNER / DUE DATE

GO / NO-GO AUTHORITY / DATE

Bring the real workflow, source and policy revisions, representative cases, existing review process, and the person who owns the decision.