Human-in-the-Loop Model

Reusable LPM Knowledge Object for human review, approval, override, escalation, evidence, and AI accountability

Use this Human-in-the-Loop Model to make human review an operating design, not a vague safety phrase. The model defines where humans belong in AI-enabled work: before the work starts, while AI is producing recommendations, before actions are taken, after actions are logged, and when exceptions or drift appear. It helps companies scale AI without pretending that a human is accountable simply because someone was nearby, copied on an email, or able to theoretically intervene.

Built for three company profiles: 500+ employees, 5,000+ employees, and 10,000+ employees. Use it as a website artifact, AI review design model, governance worksheet, and future Lapemo guided skill.

Metadata

Metadata	Value
Object type	LPM Knowledge Object
Primary LPM layer	AI Amplification
Connected layers	Ownership Map, Decision Architecture, Communication Architecture, Information Ecology, Platform Structure, Governance Architecture, Control Map, AI Governance Checklist, Agent Accountability Checklist
Primary use	Define where humans must review, approve, override, audit, intervene, or stop AI-assisted work before AI changes decisions, records, communications, workflows, controls, or outcomes.
Website use	Downloadable model, AI governance artifact, design partner worksheet, agent launch gate, JSON object for Lapemo ingestion, and future guided skill.
Version	1.0
Owner	LPM / Lapemo
Last reviewed	2026-06-24

Core principles

Principle	Meaning
Human-in-the-loop is not enough by default	A loop is only useful when the human has authority, context, evidence, time, and the ability to change or stop the outcome.
Accountability stays with humans	AI may recommend, draft, summarize, classify, route, or act, but a named human or governance forum owns the outcome.
Review depth must match impact	Low-impact tasks can use lightweight review. High-impact decisions, system actions, regulated workflows, and external communications require stronger controls.
Review must happen at the right point	A human review after a bad decision has already reached a customer, employee, regulator, or system of record is not meaningful control.
Evidence must be reviewable	Reviewers need source lineage, confidence, assumptions, model output, tool calls, exceptions, and downstream impact, not polished AI summaries alone.
The human must be empowered	A reviewer without decision rights, escalation authority, or stop rights creates false assurance.
Over-review creates theater	Putting humans everywhere creates bottlenecks and rubber-stamping. The model should place humans where judgment, authority, ethics, risk, or accountability matters.
Loops must be monitored	Override rates, review quality, delays, escalations, incidents, drift, and audit findings determine whether the loop is working.

Loop types

Loop type	Description
Human-in-the-front	A human defines intent, scope, policy, data boundary, prompt pattern, and approval rule before AI begins work.
Human-in-the-decision	AI prepares options or recommendations, but a human decision owner makes the decision.
Human-in-the-approval	AI drafts or prepares an action, but a human approves before execution, external communication, or system-of-record change.
Human-on-the-loop	AI operates within approved boundaries while humans monitor signals, exceptions, drift, quality, and risk.
Human-over-the-loop	A human or governance forum can pause, restrict, override, roll back, or retire the AI system.
Human-after-the-loop	A human audits samples, logs, outcomes, incidents, overrides, and control evidence after execution.

Required fields

Field	Definition	Required
Loop ID	Unique identifier for the human review or intervention point	Yes
AI use case or agent	Initiative, workflow, model, automation, assistant, or agent the loop applies to	Yes
Business outcome	Outcome, decision, workflow, service, control, or risk the AI supports	Yes
Impact tier	Low, moderate, high, critical, regulated, customer-facing, employee-impacting, financial, security, privacy, or control-impacting	Yes
Loop type	Human-in-the-front, human-in-the-decision, human-in-the-approval, human-on-the-loop, human-over-the-loop, or human-after-the-loop	Yes
Review trigger	Event, decision, threshold, exception, confidence score, risk condition, or workflow step that requires human involvement	Yes
Human reviewer role	Named person, role, queue, forum, control owner, decision owner, business owner, or accountable approver	Yes
Reviewer authority	Approve, reject, revise, escalate, pause, override, roll back, accept risk, or retire	Yes
Evidence required	Sources, lineage, assumptions, output, confidence, tool calls, logs, risk tier, policies, prior decisions, and control evidence reviewed by the human	Yes
Allowed AI action before review	What AI may do before a human is involved	Yes
Blocked AI action before review	What AI must not do until human review or approval happens	Yes
Decision rights link	Decision rights model, matrix, owner, or forum tied to the loop	Required for decisions
Control link	Control, monitoring signal, approval gate, logging rule, or audit requirement tied to the loop	Required for moderate and above
Escalation path	Where the reviewer sends uncertainty, conflict, failed evidence, policy exception, or unsafe output	Yes
SLA or review window	Required response time for the loop to be useful	Required for operational workflows
Override rule	When a human may override AI and what must be logged	Yes
Stop or pause rule	Who can stop the AI process and under what conditions	Required for high and above
Evidence retention rule	What review evidence is retained, where it lives, and how long it remains valid	Yes
Monitoring signals	Review backlog, approval rate, rejection rate, override rate, escalations, drift, incidents, and value impact	Yes
Review cadence	Weekly, monthly, quarterly, release-based,	Yes

	incident-based, source-change based, model-change based, or policy-change based	

Model components

Component	Description
Trigger	The condition that activates human involvement, such as low confidence, high risk, external communication, system update, policy exception, or material decision.
Reviewer	The human role with enough context and authority to judge the AI output or action.
Evidence package	The source material, lineage, assumptions, prior decisions, policies, confidence, logs, and generated output available for review.
Authority boundary	What the reviewer can approve, reject, revise, escalate, pause, override, or retire.
System action rule	What AI can and cannot do before human approval, especially for write access, external communication, commitments, or control-impacting actions.
Escalation route	The route when the reviewer cannot approve, evidence is weak, authority is unclear, or risk exceeds tolerance.
Audit trail	The log of what AI proposed, what evidence was used, who reviewed, what decision was made, and what changed downstream.
Feedback loop	How review outcomes improve prompts, sources, controls, training, workflow design, access, and governance rules.

Human-in-the-loop checklist

1. Use case and impact classification

Checklist item	Status	Guidance
AI use case or agent is named	Yes / No / Partial	The loop must attach to a real AI capability, not a broad program label.
Business outcome is defined	Yes / No / Partial	Review design should match the outcome AI is affecting.
Impact tier is assigned	Yes / No / Partial	Classify by the highest impact the AI can create, not the average task.
Regulatory, customer, employee, financial, privacy, security, and control impacts are identified	Yes / No / Partial	Higher-impact categories require stronger human authority and evidence.

2. Loop placement and trigger design

Checklist item	Status	Guidance
Loop type is selected	Yes / No / Partial	Front, decision, approval, on, over, or after the loop.
Review trigger is explicit	Yes / No / Partial	Define the exact condition that requires human review.
Human review happens before irreversible or external action	Yes / No / Partial	Approval after harm occurs is not a control.
Escalation trigger is documented	Yes / No / Partial	Define when the reviewer must escalate rather than decide.

3. Human authority and accountability

Checklist item	Status	Guidance
Reviewer role is named	Yes / No / Partial	A generic human reviewer is not enough.
Reviewer has decision rights	Yes / No / Partial	The reviewer must have authority to approve, reject, revise, escalate, or stop.
Accountable business owner is named	Yes / No / Partial	The reviewer may not always be the accountable owner, but ownership must be clear.

Control owner is named for higher-risk loops	Yes / No / Partial	Human review must connect to controls and evidence, not just personal judgment.

4. Evidence and context requirements

Checklist item	Status	Guidance
Evidence package is defined	Yes / No / Partial	Sources, assumptions, confidence, lineage, prior decisions, policies, and output must be reviewable.
AI summary is not treated as sole evidence	Yes / No / Partial	Reviewers need access to underlying source material when impact requires it.
Freshness and source-of-truth rules are defined	Yes / No / Partial	Human review is weak if the evidence is stale or unofficial.
Evidence retention rule is defined	Yes / No / Partial	Log enough to replay why the human approved, rejected, revised, or escalated.

5. AI action boundary

Checklist item	Status	Guidance
Allowed AI actions before review are listed	Yes / No / Partial	Define whether AI may retrieve, summarize, draft, classify, score, route, update, trigger, or communicate.
Blocked AI actions before review are listed	Yes / No / Partial	Block external commitments, regulated decisions, control bypass, system updates, and sensitive communication when needed.
Write access requires approval rule	Yes / No / Partial	Any system-of-record update must have explicit approval or approved automation boundary.
Customer or employee-impacting actions require stronger review	Yes / No / Partial	Human review must occur before high-impact communications or outcomes.

6. Controls, monitoring, and auditability

Checklist item	Status	Guidance
Control link is documented	Yes / No / Partial	Tie the loop to preventive, detective, corrective, access, or audit controls.
Monitoring signals are defined	Yes / No / Partial	Track override rate, rejection rate, escalation rate, backlog, incidents, drift, and quality.
Stop, pause, rollback, or override path is documented	Yes / No / Partial	The human must be able to intervene when AI behavior becomes unsafe or unreliable.
Review quality is assessed	Yes / No / Partial	A rubber-stamp loop is worse than no loop because it creates false confidence.

7. Lifecycle and improvement

Checklist item	Status	Guidance
Loop review cadence is defined	Yes / No / Partial	Review design must change as AI, data, policy, tools, and workflows change.
Review outcomes feed improvement	Yes / No / Partial	Use human decisions to improve prompts, sources, controls, training, and routing.
Retirement rule is defined	Yes / No / Partial	Remove loops that become unnecessary, ineffective, slow, or automated with sufficient controls.
Supersession rule is defined	Yes / No / Partial	When a new model, source, workflow, control, or policy changes the loop, supersede the old version.

Version for 500+ employee company

Dimension	Recommended pattern
Design intent	Create lightweight human review rules before AI tools move from assistant use to workflow, communication, decision, or system-action use.
Minimum scope	Document owner, use case, impact tier, trigger, reviewer, evidence, allowed actions, blocked actions, escalation, and review cadence.
Operating pattern	Small AI governance group or transformation owner defines common loop patterns and reviews higher-risk use cases.
AI focus	Prevent teams from relying on informal review, chat-based approvals, or unlogged human judgment.
Governance need	Connect to AI governance checklist, agent accountability checklist, decision rights model, evidence checklist, and control map.
Red flags	A human is said to be in the loop but no one knows what they review, what evidence they see, or what authority they have.

Version for 5,000+ employee company

Dimension	Recommended pattern
Design intent	Create repeatable loop patterns across business units, data domains, platforms, workflows, and governance forums.
Minimum scope	Add role-based review tiers, evidence packages, approval SLAs, escalation routing, control links, logs, and monitoring signals.
Operating pattern	Business units apply standard loop models while enterprise governance reviews high-impact, regulated, customer-facing, employee-impacting, or system-action AI.
AI focus	Prevent inconsistent human review rules across teams, duplicated approvals, weak evidence, and review bottlenecks.
Governance need	Connect to source-of-truth map, data lineage map, integration map, workflow inventory, risk acceptance register, and decision log.
Red flags	Reviewers are overloaded, approvals are rubber-stamped, or AI can act faster than humans can govern exceptions.

Version for 10,000+ employee company

Dimension	Recommended pattern
Design intent	Create enterprise-grade human accountability across regions, regulated functions, shared platforms, vendors, agents, and multi-agent chains.
Minimum scope	Full traceability across use case, owner, decision, workflow, evidence, sources, tools, controls, risk, approvals, incidents, and lifecycle state.
Operating pattern	Central AI governance defines standards and risk tiers while federated owners manage loop execution, evidence, escalation, monitoring, and improvement.
AI focus	Prioritize high-risk loops for system-action agents, external communication, regulated decisions, model-driven prioritization, employee-impacting outcomes, and customer-impacting workflows.
Governance need	Integrate with GRC, IAM, model risk, privacy, legal, security, data governance, enterprise architecture, internal audit, and executive control-plane reporting.
Red flags	A chain of agents and humans makes it unclear who reviewed, who approved, what evidence was used, and who owned the outcome.

Scoring logic

Dimension	Score	What good looks like
Impact classification	0-5	AI use case is classified by strongest possible business, customer, employee, financial, regulatory, security, privacy, and control impact.
Loop placement	0-5	Human involvement occurs at the right point before material decisions, system actions, external communication, or irreversible outcomes.
Reviewer authority	0-5	The reviewer has clear authority to approve,

		reject, revise, escalate, pause, override, or stop.
Ownership clarity	0-5	Business owner, reviewer, decision owner, technical owner, and control owner are clear where needed.
Evidence quality	0-5	Reviewer receives source lineage, freshness, confidence, assumptions, policy, prior decisions, logs, and AI output appropriate to impact.
Action boundary	0-5	AI allowed and blocked actions before review are explicit, especially for write access, commitments, and external communication.
Control coverage	0-5	Loop is tied to controls, approvals, logs, monitoring, exception handling, and auditability.
Escalation readiness	0-5	Uncertainty, policy exceptions, low confidence, high risk, and failed evidence have a defined escalation path.
Monitoring discipline	0-5	Backlog, delays, approval rate, rejection rate, override rate, incidents, drift, review quality, and value impact are monitored.
Lifecycle governance	0-5	Loop has review cadence, change triggers, retirement rule, supersession rule, and improvement path.

Suggested readiness score: average the ten scores, then classify 0-1.9 as Loop theater / unmanaged human review, 2.0-3.4 as Documented but weak intervention model, 3.5-4.4 as Governed human-in-the-loop model, and 4.5-5.0 as Enterprise-ready human accountability model.

AI prompts

- Classify this AI use case by impact tier, strongest possible action, review trigger, reviewer authority, evidence needs, control needs, and escalation route.
- Review this human-in-the-loop design and identify missing owners, weak triggers, unclear reviewer authority, insufficient evidence, missing controls, and rubber-stamp risk.
- Given this AI workflow, propose the correct loop type: human-in-the-front, human-in-the-decision, human-in-the-approval, human-on-the-loop, human-over-the-loop, or human-after-the-loop.
- Generate allowed AI actions before review, blocked AI actions before review, human approval rules, escalation triggers, evidence package, and monitoring signals.
- Score this loop from 0 to 5 across impact classification, loop placement, reviewer authority, ownership clarity, evidence quality, action boundary, control coverage, escalation readiness, monitoring discipline, and lifecycle governance.
- Convert this human-in-the-loop model into Lapemo objects for AI initiative, agent, owner, reviewer, decision, source, workflow, control, evidence, risk, escalation, monitoring signal, and review cadence.

Validation rules

- Every material AI use case must have an explicit human accountability model before production use.
- Every loop must define trigger, reviewer role, reviewer authority, evidence package, allowed AI actions, blocked AI actions, escalation path, and monitoring signal.
- A named accountable business owner must exist even when review is performed by a queue, forum, control owner, or delegate.
- High-impact, critical, regulated, customer-facing, employee-impacting, financial, security, privacy, or control-impacting AI must have human approval or governance-approved automation boundaries before material action.
- Human review may not rely only on AI-generated summary when the impact tier requires source evidence or lineage.
- Reviewers must have authority to approve, reject, revise, escalate, pause, override, roll back, or stop according to risk tier.
- System-of-record updates, external commitments, regulated decisions, control exceptions, and sensitive communications require explicit approval rules.
- Every loop above low impact must link to at least one control, log, evidence requirement, and escalation path.
- Rubber-stamp loops must be redesigned, automated with stronger controls, or removed.
- Loop design must be reviewed when sources, model, vendor, workflow, policy, controls, ownership, impact tier, or system access changes.

Lapemo ingestion mapping

Lapemo object	Fields / entities	Use
AI Initiative	initiative_id, name, purpose, impact_tier, lifecycle_status	Connects loop to active AI work.
Agent	agent_id, agent_type, allowed_actions, blocked_actions, tool_access	Defines whether loop applies to an agent or agent chain.
Owner	business_owner, reviewer_role, technical_owner, control_owner, decision_owner	Preserves accountability across humans and systems.
Decision	decision_type, decision_rights_link, approval_rule, supersession_rule	Connects AI output to decision authority.
Workflow	workflow_id, trigger, handoff, SLA, system_action_rule	Places the loop inside real operating flow.
Evidence	source_lineage, confidence, assumptions, logs, retained_evidence	Defines what the human reviews and what is stored.
Control	control_id, control_type, monitoring_signal, exception_rule, audit_rule	Connects loop to governance control.
Risk	risk_tier, risk_acceptance_link, escalation_path, residual_risk	Determines governance route and acceptance needs.
Monitoring Signal	approval_rate, rejection_rate, override_rate, backlog, incidents, drift	Measures whether the loop is working.
Knowledge Object	version, owner, review_cadence, last_reviewed, superseded_by	Keeps the model reusable and current.

Website positioning

Use this artifact as a downloadable AI Amplification model and as a future guided skill inside Lapemo. The website version should make a strong claim: human-in-the-loop is not a checkbox. It is the operating model that determines when human judgment, authority, evidence, control, and accountability are required before AI work becomes an enterprise outcome.

Reusable knowledge object structure

Layer	Reusable asset
Human guide	Plain-language explanation of where humans belong in AI-enabled work.
Template	Downloadable DOCX/PDF/Markdown model for workshops and governance reviews.
Machine schema	JSON structure for ingestion into Lapemo and future automation.
Guided skill	AI-assisted workflow that classifies AI use cases, recommends loop placement, scores readiness, and routes governance.