

Human Oversight Record

Review and approval of sensitive AI outputs or actions

Core oversight commitment

KOMGIS will not treat an AI model as the final accountable decision-maker for sensitive outputs or actions. Where review is required, the system must present sufficient evidence to an authorized human reviewer and must block the sensitive action until the required approval is recorded. Reviewers must be able to reject, correct, override, escalate, stop or disable the affected workflow.

Record owner: AI Governance / Technical Leadership

Version: 1.0

Effective date: 8 July 2026

Classification: Internal Governance Record / Controlled Template

Document Control

Field	Value
Document title	KOMGIS Human Oversight Record
Document owner	AI Governance / Technical Leadership
Version	1.0
Effective date	8 July 2026
Review frequency	At least annually and after defined trigger events
Approved by	[Executive approver / role]
Classification	Internal Governance Record / Controlled Template
Applies to	KOMGIS employees, contractors, AI systems, client implementations and approved third parties involved in AI review, approval or sensitive actions
Related parent policy	KOMGIS Responsible AI and AI Management Policy
Record status	Master oversight procedure and template set; completed instances form auditable governance records

Status and governance note

This document defines KOMGIS's operational record for designing, performing, documenting and auditing human oversight of AI outputs and actions. It is intended to be used with the AI System Inventory, AI Risk Assessment / Impact Assessment, AI Security Policy, Data Privacy Policy, Model Provider Disclosure / Vendor List, Evaluation and Testing Records and AI Incident Response Plan. It does not by itself determine legal classification or certify compliance with a particular regulatory regime. The applicable oversight arrangement must be assessed for each AI system, intended purpose, deployment context, client requirement and risk classification.

Mandatory approval rule

Where a sensitive action is designated as approval-gated, the application or workflow must prevent execution until the required approval is issued by an authorized reviewer. A model response, confidence score, prompt instruction or agent retry must never substitute for the required human approval.

Contents

1. Purpose and Objectives
2. Scope
3. Governance Basis and Relationship to Other Records
4. Definitions
5. Human Oversight Principles
6. Roles and Accountability
7. When Human Oversight Is Required
8. Oversight Level Classification
9. Sensitive Output and Action Categories
10. Oversight Model Design
11. Reviewer Eligibility, Competence and Authority
12. Evidence Available to the Reviewer
13. Standard Review and Approval Workflow
14. Review Decision Criteria
15. Decision Codes, Rejection and Escalation
16. Material Edits and Corrections
17. Sensitive Action Approval Gates
18. RAG and Document Intelligence Oversight
19. Summarization, Extraction and Classification Oversight
20. AI Agents and Tool-Use Oversight
21. External Communications and Client-Facing Outputs
22. Sensitive-Domain and Important-Decision Support
23. Automation Bias and Reviewer Workload Safeguards
24. Stop, Override, Disable and Emergency Intervention
25. Escalation and Second Review
26. Logging, Traceability and Audit Evidence
27. Privacy, Confidentiality and Retention
28. Monitoring Oversight Effectiveness
29. Sampling and Retrospective Quality Review
30. Incident Response and Nonconformity Linkage
31. Change Control and Revalidation
32. Client-Specific Oversight Requirements
33. Audit, Management Review and Continual Improvement
34. Record Ownership, Review and Enforcement
- Appendix A. Oversight Level Matrix
- Appendix B. Human Oversight Plan Template
- Appendix C. Sensitive Output / Action Review Record
- Appendix D. Reviewer Checklist
- Appendix E. AI Agent Action Approval Record
- Appendix F. RAG / Document-Grounded Review Record
- Appendix G. Escalation and Second Review Log
- Appendix H. Reviewer Competence and Authorization Record
- Appendix I. Oversight Metrics and Threshold Register
- Appendix J. Decision and Reason Codes
- Appendix K. Worked Oversight Examples

1. Purpose and Objectives

This record defines the minimum process and evidence KOMGIS uses to demonstrate meaningful human review of AI-generated outputs, recommendations, tool calls and actions that could create material client, individual, legal, financial, privacy, security, operational or reputational impact.

The objectives are to ensure that:

- human accountability is preserved for sensitive business outcomes;
- review occurs at the correct stage, before an irreversible or material action where required;
- reviewers receive enough context, source evidence and system information to make an independent decision;
- reviewers can approve, reject, correct, override, stop or escalate the AI output or proposed action;
- approval gates are enforced in application logic and are not left to prompt wording or voluntary user behavior;
- the reviewer, decision, reviewed output, material edits, action and timestamp are traceable;
- oversight effectiveness is monitored and improved using rejection, edit, escalation, incident and bypass-attempt evidence.

2. Scope

This record applies to production and client-facing AI systems developed, configured, integrated, deployed or operated by KOMGIS when human oversight is required by KOMGIS policy, risk assessment, contractual requirement, system design or applicable law.

- AI assistants and chatbots that generate client-facing or material business content;
- RAG and enterprise document intelligence systems;
- document summarization, extraction, classification and report-generation systems;
- AI agents and tool-using systems, especially those with write or external-action capabilities;
- workflow automation and decision-support systems;
- AI-generated external email, notices, reports, forms, approvals or submissions;
- AI features used in sensitive domains or important decision processes;
- human reviewers, system owners, technical owners, project managers and approved client reviewers.

Low-risk internal assistance may rely on normal user judgment rather than a formal completed oversight record.

However, a formal Human Oversight Plan is required whenever the AI System Inventory or risk assessment identifies a defined review or approval step, and completed review records are required for designated sensitive outputs or actions.

3. Governance Basis and Relationship to Other Records

Related governance record	Human oversight relationship
KOMGIS Responsible AI and AI Management Policy	Sets the human accountability, human oversight, sensitive-action approval and traceability requirements.
AI System Inventory	Identifies whether the system requires human review, the approval model, responsible owners, users, tools and operational status.
AI Risk Assessment / Impact Assessment	Identifies harms, affected parties, automation-bias risk, reversibility, reviewer competence needs and residual risk.
AI Security Policy	Requires server-side authorization, least privilege, separation of read/write capabilities and protection against approval bypass.
Data Privacy Policy	Controls personal and client data visible to reviewers and retained in review records.
Model Provider Disclosure / Vendor List	Provides provider context where model limitations, retention or processing arrangements affect the review design.
Evaluation and Testing Records	Provide evidence that review workflows, approval gates, citations, tool permissions and failure behavior have been tested.
AI Incident Response Plan	Defines containment and escalation when oversight fails, a reviewer detects harmful behavior or an action bypasses approval.

The Human Oversight Record does not duplicate system-specific risk or test evidence. It records the oversight design and the human decision at the point where KOMGIS requires human judgment or authorization.

4. Definitions

Term	Record meaning
Human oversight	A control in which an authorized natural person meaningfully reviews AI output, recommendation, evidence or proposed action and has authority to influence the outcome.
Human reviewer	A person assigned to inspect evidence and approve, reject, correct, override, stop or escalate an AI output or action.
Sensitive output	AI-generated content or recommendation whose use may materially affect a client, individual, regulated process, external communication or important business outcome.
Sensitive action	An external, irreversible, high-impact, regulated or materially consequential action requiring defined human approval.
Approval gate	An application or workflow control that blocks a designated action until the required reviewer decision is recorded.
Pre-action review	Review that must be completed before an AI-proposed action is executed or an output is used in a sensitive process.
Post-action review	Retrospective review of completed outputs or actions for monitoring, quality assurance or sampling. It is not a substitute for a required pre-action approval.
Material edit	A human change that alters facts, reasoning, recommendation, recipient, scope, amount, legal meaning, action parameters or other business-significant content.
Override	A human decision to disregard or replace the AI output or proposed action.
Escalation	Transfer of a case to a higher-authority, second-line, specialist, security, privacy, legal/compliance or management reviewer.
Automation bias	A tendency to accept AI output because it appears confident, polished or system-generated rather than because the evidence supports it.
Decision code	A standardized status used to record approval, rejection, correction, escalation, cancellation or other review outcomes.

5. Human Oversight Principles

5.1 Human accountability

The accountable system owner and authorized human process owner remain responsible for sensitive business outcomes. Responsibility must not be assigned solely to the model or model provider.

5.2 Meaningful review

A click-through confirmation without access to relevant evidence is not meaningful oversight. The reviewer must have enough information, time, competence and authority to make an independent decision.

5.3 Review before consequence

Where an action is external, irreversible or materially consequential, review occurs before the consequence unless a specifically assessed and approved automation model documents a lawful exception.

5.4 Ability to disagree

The review interface and process must permit rejection, correction, override and escalation. A reviewer must not be forced to approve merely to continue the workflow.

5.5 Evidence over confidence

Model confidence language, polished prose or internal scores do not replace source verification, business rules or professional judgment.

5.6 Proportionate oversight

The level of review increases with autonomy, data sensitivity, affected persons, scale, reversibility, domain sensitivity and potential harm.

5.7 Traceability

Sensitive decisions and actions must be linked to the reviewed output, reviewer, decision, material edits, action and relevant case or workflow identifier.

5.8 Safe intervention

Reviewers and authorized operators must be able to stop or disable the affected workflow when anomalous, unsafe or unauthorized behavior is detected.

6. Roles and Accountability

Role	Minimum responsibilities for human oversight
Executive leadership	Approve policy direction; review critical oversight models, material exceptions and severe oversight failures; support corrective action.
AI Governance / Technical Leadership	Maintain this record and oversight methodology; approve oversight levels; coordinate governance review and audit.
AI system owner	Define the oversight model; ensure reviewers and approval criteria exist; accept residual risk; review effectiveness and changes.
Technical owner	Implement approval gates, evidence presentation, audit logging, reviewer authorization, stop/disable controls and monitoring.
Security / privacy owner	Review access, client isolation, data exposure, sensitive content, approval bypass and reviewer access to governed data.
Project manager / client owner	Capture client-specific review requirements; coordinate reviewer assignments and client communication.
Human reviewer	Inspect relevant evidence; apply criteria independently; approve, reject, correct, override or escalate; report suspicious behavior.
Second reviewer / specialist	Review escalated or dual-approval cases; provide specialist judgment for legal, security, privacy, financial or domain-sensitive matters.
Support / operations	Preserve review evidence, route incidents, assist safe-stop or disable procedures and avoid modifying audit records.

Named accountability rule

Every production AI system with formal human oversight must identify the AI system owner, technical owner, reviewer role or reviewer pool, escalation owner and approval authority. A generic label such as “business user” is insufficient for a high or critical oversight requirement.

7. When Human Oversight Is Required

Formal human oversight is required when identified by the AI System Inventory, risk assessment, client requirement, system design or applicable obligation. The system owner shall apply the most protective requirement where multiple requirements overlap.

Mandatory or normally required triggers include:

- sending external email, customer communication, notices or reports where content can materially affect a client or individual;
- updating official client, legal, HR, insurance, financial, compliance or other important records;
- triggering payment, financial change, credit, pricing, refund or other monetary consequence;
- sharing files, records or client data externally;
- submitting forms, applications, approvals or regulatory/contractual responses;
- deleting, closing or materially changing important records, cases or complaints;
- using AI recommendations in important decisions affecting rights, access, money, health, employment, education, insurance, safety or essential services;
- accepting a high-impact extraction, classification or summary as an official business record without independent validation;
- agent actions with write tools, elevated credentials or difficult-to-reverse effects;
- situations where the risk assessment identifies meaningful automation-bias, fairness, hallucination or evidence-verification risk.

Human oversight may also be implemented as sampling, exception review or post-action quality review for lower-risk outputs. Such review must be clearly distinguished from mandatory pre-action approval.

8. Oversight Level Classification

KOMGIS uses four oversight levels to convert risk and action characteristics into an operational review model. The level is recorded in the Human Oversight Plan and linked to the AI System Inventory and risk assessment.

Level	Typical use	Minimum oversight
O0 - Normal user judgment	Low-risk internal assistance, no sensitive action, easily reversible errors.	No formal approval record normally required. Users remain responsible for normal verification and incident reporting.
O1 - Defined review / sampling	Client-facing support, RAG answers, summaries or extractions with limited consequence.	Defined reviewer guidance; evidence available; targeted review or sampling; correction and escalation; metrics monitored.
O2 - Mandatory single approval	Sensitive output, material external communication or controlled write action.	Pre-action review by one authorized reviewer; system blocks action until approval; complete review record required.
O3 - Enhanced / dual approval	High or critical impact, high autonomy, difficult-to-reverse action, special-domain or severe consequence.	Pre-action review; specialist or second reviewer where defined; explicit approval chain; enhanced logging; stop/disable capability; management or control-owner acceptance.

The oversight level shall consider the most consequential reasonably foreseeable scenario, not the average expected case. Increasing automation, adding write tools, removing an approval step or expanding to a sensitive domain requires reassessment.

9. Sensitive Output and Action Categories

Category	Examples	Default review expectation
External communication	Client email, customer notice, formal response, published statement.	O2 unless low consequence and specifically approved for O1.
Official record change	Case status, claim record, HR record, compliance status, contract metadata.	O2; O3 for high-impact or difficult-to-reverse changes.
Financial action	Payment, refund, price change, invoice approval, monetary adjustment.	O2 or O3 according to amount, authority and reversibility.
Data sharing / export	Sharing files, bulk export, external transfer, permissions change.	O2; O3 for sensitive, large-scale or cross-client exposure risk.

Category	Examples	Default review expectation
Deletion / closure	Delete record, close case, close complaint, revoke access.	O2; O3 where recovery is limited or rights are materially affected.
Important decision support	Employment, insurance, eligibility, health, education, credit, safety.	O3 or domain-specific formal review; AI recommendation is not treated as final by default.
High-impact document output	Legal summary, compliance assessment, policy interpretation, formal report.	O2 or O3 based on reliance and consequence.
Agent tool action	API write, form submission, database update, email send, ticket closure.	Approval level based on tool scope and consequence; write actions normally O2+.

10. Oversight Model Design

The system owner and technical owner shall define the oversight model before production use. The design must be specific enough to implement and test.

- which outputs or actions trigger review;
- whether review is pre-action, post-action, sampling, exception-based or continuous supervisory review;
- the oversight level and required number or type of reviewers;
- who is authorized to approve and how reviewer authorization is enforced;
- what evidence and source material the reviewer can inspect;
- the approval, rejection, correction, escalation and cancellation options;
- what the system does while approval is pending;
- how expired, abandoned or conflicting approvals are handled;
- whether approval is one-time, action-specific or valid for a defined group of actions;
- how the final action is bound to the approved content and parameters;
- the stop, disable, rollback or safe-state mechanism;
- logging, retention, monitoring and audit requirements.

Approval binding rule

Approval must apply to the exact output or action parameters reviewed. If a material field changes after approval - such as recipient, attachment, amount, record, tool, action type or material content - the prior approval is invalid unless the oversight design explicitly defines and validates a safe non-material change rule.

11. Reviewer Eligibility, Competence and Authority

A reviewer must be competent for the business context and authorized to make the required decision. The system owner shall document reviewer eligibility in the Human Oversight Plan.

Competence area	Minimum expectation
System understanding	Understands intended purpose, relevant capabilities, known limitations and common failure modes.
Evidence verification	Can inspect and interpret source documents, data or business records relevant to the review.
Process knowledge	Understands the business process, decision criteria and consequences of approval or rejection.
AI literacy	Recognizes hallucination, unsupported claims, misleading confidence, prompt injection indicators and automation bias.
Authority	Has organizational authority to approve, reject, correct, override, stop or escalate the reviewed matter.
Security / privacy awareness	Understands confidentiality, client isolation and the limits on copying or exposing review data.
Escalation judgment	Knows when uncertainty, conflict, high impact or suspected incident requires specialist or higher-level review.

For O3 systems, KOMGIS should document role-specific training or demonstrated competence. A reviewer who is the direct author of a material manual change may be restricted from providing the second approval where independence is required.

12. Evidence Available to the Reviewer

The review interface or review package shall present evidence proportionate to risk. It should minimize unnecessary data while avoiding a “blind approval” design.

Evidence item	When applicable
AI output or proposed action	Always. The exact content and parameters subject to approval must be visible.
User request / business instruction	Where needed to understand the intended task or scope.
Source citations and retrieved documents	For RAG, document-grounded answers and material summaries.
Original source record / extracted fields	For extraction, classification and record-update workflows.
Tool name, action type and target	For agent or automated tool calls.
Recipient, attachment and sharing scope	For external communication or data transfer.
Model / workflow version	For traceability, high-risk review, incident investigation or controlled releases.
Validation and policy checks	Schema status, business-rule status, permission status and relevant warnings.
Prior review history	Where repeated review, escalation or conflicting decisions are relevant.
Material risk warning	Where the risk assessment identifies a reviewer-specific caution, limitation or prohibited reliance.

The reviewer should not receive hidden chain-of-thought or internal model reasoning as a substitute for evidence. The reviewer needs the business-relevant basis, source material, outputs, action parameters and control results needed to make an independent decision.

13. Standard Review and Approval Workflow

1. The AI system produces an output, recommendation or proposed action and assigns the relevant workflow or case identifier.
2. Application logic determines whether the item is review-gated and identifies the required oversight level.
3. For a gated action, the action is placed in a pending state and execution is blocked.
4. The review package is created with the exact output or action parameters, supporting evidence and required warnings.
5. An authorized reviewer is assigned or selects the item from an authorized review queue.
6. The reviewer verifies identity of the target case or workflow, inspects evidence and applies the defined review criteria.
7. The reviewer records an approval, rejection, correction, override, escalation or cancellation decision using the appropriate decision and reason codes.
8. Where edits are made, the final reviewed version and material edits are preserved or version-linked.
9. The system validates that the reviewer is authorized and that the approved output or action still matches the reviewed version.
10. Only after successful validation does the system execute the approved action or release the approved output.
11. The final action status, reviewer decision, timestamp and audit reference are recorded.
12. Monitoring and incident controls evaluate rejection trends, bypass attempts, unusual actions and other risk indicators.

14. Review Decision Criteria

Review criteria must be defined for the actual intended purpose. The following baseline criteria apply where relevant:

Criterion	Reviewer question
-----------	-------------------

Criterion	Reviewer question
Identity and scope	Is this the correct client, workspace, person, case, record, recipient and task?
Purpose	Is the output or action within the approved intended purpose and business instruction?
Evidence	Do the cited or source records support the material statements or extracted fields?
Accuracy	Are key facts, dates, amounts, names, classifications and action parameters correct?
Completeness	Is material context missing in a way that could mislead or change the outcome?
Permissions	Is the data, document, tool and target authorized for this user and client context?
Confidentiality	Does the output avoid unnecessary or unauthorized personal, confidential or cross-client information?
Business rules	Does the item comply with required business rules, thresholds and process controls?
Tone and communication	For external content, is the message appropriate, clear and free from unsupported commitments?
Reversibility / consequence	Is the consequence understood and is any higher approval level required?
Uncertainty	Is uncertainty material enough to require correction, additional evidence or escalation?
Anomalous behavior	Is there any sign of prompt injection, wrong-client retrieval, tool misuse or unexpected AI behavior?

15. Decision Codes, Rejection and Escalation

Code	Decision	Operational meaning
APP	Approved	Reviewed item may proceed exactly as approved.
APP-E	Approved after edit	Reviewer made and approved material edits; final reviewed version may proceed.
REJ	Rejected	Item must not proceed. A new output or action requires a new review cycle.
OVR	Human override	Reviewer disregards the AI recommendation and records the human-selected outcome.
ESC	Escalated	Item is routed to a second reviewer, specialist or control owner. Action remains blocked.
CAN	Cancelled	Business task or proposed action is no longer required; no execution occurs.
INC	Incident suspected	Review stops and the incident route is initiated. Action remains blocked.
EXP	Approval expired / invalidated	Prior approval no longer matches the reviewed version, time window or action state.

A rejection or escalation should include a reason code and, where useful, short reviewer notes. Free-text explanations should not be used to store unnecessary sensitive data.

16. Material Edits and Corrections

The record must distinguish between an AI output approved as generated and an output approved only after human correction. Material edits are an important oversight-effectiveness signal and shall be traceable for O2 and O3 workflows.

- factual correction or unsupported claim removal;
- change to recipient, person, case, account, client or record;
- change to amount, date, deadline, classification or status;
- change to recommendation, conclusion, outcome or decision basis;
- addition or removal of material qualification, exception or warning;
- change to tool, target, action type, permissions or external-sharing scope;
- substantial rewrite of client-facing or regulated communication.

Pure formatting, spelling or non-material stylistic edits may be recorded as non-material where the Human Oversight Plan defines that treatment. The final version actually used or executed must be identifiable.

17. Sensitive Action Approval Gates

Approval gates are technical controls. For O2 and O3 actions, the application shall enforce the gate server-side or in the trusted workflow control layer.

Control requirement	Expected implementation behavior
Pending state	Sensitive action remains non-executable until the required approval record exists.
Reviewer authorization	Only an authorized reviewer role or specifically delegated reviewer can approve.
Action fingerprint / version binding	Approval is associated with the exact action parameters or reviewed output version.
Revalidation before execution	System confirms target, action and reviewed version have not materially changed.
Single-use or scoped approval	Approval cannot be reused for unrelated actions unless the approved design explicitly permits a defined batch or bounded scope.
Agent bypass resistance	Agent cannot bypass approval by retrying, rephrasing, choosing another tool or splitting one sensitive action into smaller calls.
Expiry handling	Time-limited approvals expire safely and require fresh review when the configured validity period ends.
Audit evidence	Decision, reviewer, timestamp, action, final status and relevant identifiers are logged.
Safe failure	Authorization, logging or approval-service failure causes the action to remain blocked rather than fail open.

No prompt-only approval gate

A system prompt stating “ask the user before sending” is not sufficient for a sensitive action. The tool or workflow must enforce the approval state outside the model and must independently authorize the requested action.

18. RAG and Document Intelligence Oversight

For KOMGIS document intelligence platforms, human oversight must work with permission-aware retrieval and source traceability. Reviewers should be able to verify the answer against the authorized source material used by the system.

- display source document and source location or chunk reference for material claims where the system is intended to answer from enterprise documents;
- permit the reviewer to inspect the underlying source in the same client and permission context;
- show when no sufficient source was retrieved or when citation coverage is incomplete;
- treat unsupported material claims as rejection, correction or escalation triggers according to the use case;
- record wrong-client, unauthorized-source or suspicious retrieval as a security or AI incident concern;
- do not use human review as a substitute for tenant, user or document-level authorization before retrieval.

For high-impact document-grounded outputs, the Human Oversight Plan should define whether every material claim must be source-supported, whether certain source types are authoritative, and whether conflicting sources require escalation.

19. Summarization, Extraction and Classification Oversight

AI task	Reviewer focus	Possible approval rule
Summarization	Material facts, omissions, dates, obligations, exclusions, conclusions and source support.	Approve only after checking defined critical facts or cited passages.
Structured extraction	Field value, source location, format, required fields and validation failures.	Critical fields require source comparison; low-confidence or failed validation routes to review.

AI task	Reviewer focus	Possible approval rule
Classification	Correct class, consequences of misclassification, ambiguous cases and class-specific thresholds.	Sensitive classes require human confirmation; uncertain cases escalate.
Report generation	Source basis, calculations, conclusions, disclaimers, client context and final recipient.	Material reports require O2/O3 review before release.
Record update proposal	Target record, changed fields, old/new values and business authorization.	Write remains blocked until exact field changes are approved.

Reviewer instructions should identify “critical fields” or “critical facts” rather than asking the reviewer to generically “check the AI.” This makes oversight testable and supports meaningful monitoring.

20. AI Agents and Tool-Use Oversight

AI agents may propose actions but must not decide whether a defined human approval requirement applies. The trusted application or workflow determines the approval requirement based on tool, user, client, action and risk context.

Agent action type	Oversight expectation
Read-only retrieval	O0/O1 where authorized and low consequence; log and monitor unusual volume or sources.
Draft creation	O1 or O2 according to intended use; human remains responsible for final sensitive content.
External send / submit	O2 minimum; exact recipient, content and attachments reviewed before execution.
Database or system write	O2; show target, current value and proposed value. O3 for important or difficult-to-reverse records.
Payment / financial action	O2/O3 with applicable amount and authority controls outside the model.
Deletion / closure	O2/O3; show object, consequence and recovery status; dual approval where risk warrants.
Bulk action / export	O3 when scale materially increases impact; show count, scope and destination.
Elevated or administrative tool	O3 or prohibited unless specifically approved, narrowly scoped and strongly monitored.

The approval record should capture the tool, tool version or action type where relevant, target object, material parameters, client/workspace, requested action, reviewer decision and execution result.

21. External Communications and Client-Facing Outputs

AI-generated external content requires review proportionate to the likelihood that recipients will rely on it and the consequence of error. The Human Oversight Plan shall identify which communication types are approval-gated.

- verify the recipient and client context;
- verify material statements against source or business records;
- remove confidential or cross-client information;
- check for unsupported commitments, guarantees, deadlines, legal conclusions or payment instructions;
- confirm tone and clarity are appropriate for the communication type;
- review attachments and sharing scope separately from the message text;
- ensure the approved version is the version actually sent.

For routine, low-consequence messages, the system owner may approve O1 sampling or user-confirmation controls. The decision and scope must be documented in the Human Oversight Plan and supported by evaluation evidence.

22. Sensitive-Domain and Important-Decision Support

Projects involving employment, education, health, insurance, credit, essential services, legal processes, law enforcement, migration, justice, biometrics, vulnerable persons or safety require early escalation under the parent policy. Human oversight for such uses must be designed with the actual domain, KOMGIS role and applicable obligations in mind.

- AI recommendations are not treated as final decisions by default;
- reviewers must have appropriate domain competence and authority;
- material evidence, uncertainty and limitations must be visible;
- reviewers must be able to override and document a human-selected outcome;
- affected-person or client contest, correction or appeal processes must be considered where applicable;
- workload and automation-bias risks must be assessed, especially where review volume is high;
- dual or specialist review should be used where a single general reviewer cannot meaningfully evaluate the consequence.

Sensitive-domain caution

Adding a human approval button does not automatically make a sensitive AI use safe or compliant. The system must still be assessed for intended purpose, regulatory classification, risk, data governance, accuracy, security, transparency and impact on affected persons.

23. Automation Bias and Reviewer Workload Safeguards

Human oversight can fail when reviewers are overloaded, lack evidence or routinely approve AI output without independent judgment. The oversight design shall consider the conditions under which meaningful review is realistically possible.

Risk indicator	Possible safeguard
Very high approval rate with near-zero edits	Sample approved cases, interview reviewers, test evidence visibility and check whether review criteria are meaningful.
Review queue backlog	Adjust staffing, rate-limit gated automation, prioritize high-impact cases or reduce AI-generated volume.
Very short review times	Investigate click-through behavior; require critical-field confirmation or evidence interaction where proportionate.
Model confidence or persuasive wording dominates review	Hide unnecessary confidence language; emphasize source evidence and structured checks.
Repeated same reviewer / conflict of interest	Use reviewer pools, second review or separation-of-duties rules where needed.
High edit or rejection rates	Investigate model, prompt, retrieval, training or intended-purpose mismatch; do not normalize poor AI performance through permanent human correction.
Reviewer uncertainty not escalated	Simplify escalation, add specialist access and train on reason codes and incident indicators.

Oversight metrics are governance signals, not individual performance quotas. KOMGIS shall avoid incentives that pressure reviewers to approve faster at the expense of careful review or incident reporting.

24. Stop, Override, Disable and Emergency Intervention

Reviewers and authorized operational roles must know how to prevent continuing impact when they identify anomalous or unsafe behavior.

Intervention	Use
Reject / override item	Prevent the current output or action from being used or executed.
Pause workflow	Stop processing of the current case or queue while uncertainty is resolved.
Disable tool	Remove a specific action capability without disabling unrelated AI functions.
Disable AI feature	Stop the affected feature when behavior may be systemic or unsafe.
Revoke credential / integration	Prevent further access or action when credentials or integration scope may be compromised.
Rollback release	Return to a previously approved model, prompt or workflow version where supported.

Intervention	Use
Incident escalation	Preserve evidence and activate the AI Incident Response Plan for actual or suspected harmful, unauthorized or policy-violating behavior.

A reviewer must not silently correct and approve a case when the behavior suggests a systemic security, privacy, client-isolation or approval-bypass problem. Such behavior must be escalated and preserved as potential incident evidence.

25. Escalation and Second Review

The Human Oversight Plan shall define escalation criteria and the role receiving escalated cases. The action remains blocked until the escalation is resolved or the case is cancelled.

- insufficient or conflicting source evidence;
- uncertainty about client, person, case, target record or recipient;
- possible wrong-client retrieval or unauthorized data exposure;
- possible prompt injection, tool misuse or model manipulation;
- material legal, compliance, privacy, security or contractual interpretation;
- high-impact outcome outside the first reviewer's authority;
- conflict of interest or independence concern;
- repeated model failure pattern or unexpected behavior;
- proposed exception to a defined approval requirement;
- suspected incident or serious nonconformity.

For O3 systems, KOMGIS should define when dual approval is sequential, when a specialist reviewer is mandatory and whether the second reviewer sees the first reviewer's decision before making an independent assessment.

26. Logging, Traceability and Audit Evidence

O2 and O3 review records should capture the following minimum fields. O1 records may use a reduced dataset proportionate to the use case.

Record field	Purpose
Review record ID	Unique identifier for the oversight decision.
AI system / system version	Links the decision to the governed system.
Model and workflow / prompt version	Supports traceability and incident investigation where required.
Client / workspace / project	Identifies the authorized organizational context.
Case / workflow / task ID	Links the review to the business process.
Requesting user or service	Identifies the origin of the AI task or action request.
Output / action version or fingerprint	Binds approval to the reviewed item.
Source identifiers / evidence references	Supports verification without unnecessarily duplicating source content.
Reviewer and reviewer role	Demonstrates authorization and accountability.
Decision and reason code	Records disposition and supports monitoring.
Material edits	Distinguishes AI-as-generated approval from human-corrected approval.
Approval / review timestamp	Shows when oversight occurred.
Execution / release status	Shows whether and how the approved action was completed.
Escalation / incident reference	Links exceptional cases to further governance records.

Audit records shall be protected against unauthorized modification. Where technically practical, the system should preserve an immutable or append-only event trail for material approvals and action execution status.

27. Privacy, Confidentiality and Retention

Human review does not create unrestricted access to client or personal data. Reviewer access shall follow least privilege, client isolation and purpose limitation.

- show only evidence necessary for the review task;
- restrict reviewers to authorized client, workspace, project, case or record scopes;
- avoid duplicating entire source documents in the oversight record when source identifiers and authorized links are sufficient;
- do not store secrets, passwords, access tokens or private keys in reviewer notes;
- minimize free-text notes containing personal or confidential data;
- apply defined retention to review records, outputs, source references and audit logs;
- handle deletion requests and source-document deletion according to the Data Privacy Policy and applicable audit requirements.

A review record may require longer retention than a temporary AI output where the record demonstrates authorization, security, contractual performance or compliance. The retention basis and period should be documented in the Human Oversight Plan or linked retention schedule.

28. Monitoring Oversight Effectiveness

The system owner shall monitor whether oversight is functioning as an effective control, not merely whether approval records exist.

Metric / signal	Governance interpretation
Approval rate	Very high or sudden changes may indicate automation bias, changed case mix or deteriorating review quality.
Rejection rate	May reveal AI quality problems, intended-purpose mismatch or new failure patterns.
Material edit rate	Shows how often human correction is necessary before use.
Escalation rate	Indicates uncertainty, specialist needs or poorly defined review criteria.
Median and extreme review time	Very short times may indicate superficial review; very long times may indicate workload or evidence problems.
Approval bypass attempts	Security and control signal requiring investigation.
Expired / invalidated approvals	May reveal workflow friction, action drift or version-binding problems.
Post-approval incident rate	Measures whether approved outputs or actions still produce material failures.
Reviewer agreement / disagreement	Useful for dual-review or sampled re-review of sensitive cases.
Queue backlog and reviewer workload	Supports capacity planning and meaningful-review assessment.

Thresholds should be based on the use case and risk. They are recorded in Appendix I or the system monitoring plan. A threshold breach triggers investigation, not automatic blame of the reviewer or model.

29. Sampling and Retrospective Quality Review

KOMGIS may use retrospective review to evaluate oversight quality, especially for O1 workflows or to validate O2/O3 reviewer consistency. Sampling shall be risk-based and documented.

- random sample of approved outputs;
- targeted sample of specific model, workflow or release versions;
- reviewer-specific or team-specific sample used for training and process quality, not arbitrary punitive ranking;
- high-impact or unusual-action sample;
- cases with short review time, no evidence interaction or repeated approval patterns;
- post-incident sample of similar systems, clients, source types or action classes.

Retrospective findings may result in reviewer coaching, criteria clarification, workflow redesign, additional testing, model or prompt change, stronger validation, increased oversight level or incident/nonconformity action.

30. Incident Response and Nonconformity Linkage

The Human Oversight Record is a key evidence source when an AI incident or governance nonconformity occurs. Review evidence must be preserved when a reviewer identifies or may have approved a harmful, wrong, unauthorized or policy-violating AI behavior.

Oversight event	Required response
Wrong-client retrieval or cross-client content	Stop review, block action, preserve sources and logs, escalate as a security/privacy AI incident.
Approval bypass or unauthorized execution	Contain immediately, disable affected tool or workflow, preserve approval and execution events, activate incident response.
Harmful or dangerous recommendation	Reject, prevent use, escalate according to severity and domain, preserve the reviewed output and source context.
Repeated unsupported output	Document pattern, notify system owner, review similar outputs and evaluation evidence, consider nonconformity or incident route.
Reviewer approved incorrect output	Contain impact, investigate AI, evidence presentation, criteria, workload and human factors; do not assume individual error is the sole root cause.
Required review step omitted	Block or suspend affected workflow where necessary; assess affected actions; record nonconformity and investigate control failure.

Serious oversight failures must not be hidden through silent correction. Incident, risk, testing, training, monitoring and oversight design records shall be updated as required by the corrective-action process.

31. Change Control and Revalidation

The oversight design is a controlled part of the AI system. The system owner and technical owner shall review the Human Oversight Plan when a material change affects risk, evidence, reviewer capability or approval behavior.

- new intended purpose, user group or client context;
- new model provider or materially different model behavior;
- new RAG data source or permission model;
- new write tool, integration or external action;
- increase in autonomy or action scale;
- removal, reduction or automation of a human review step;
- change to reviewer criteria, reviewer role or approval authority;
- change to evidence shown to reviewers;
- material increase in rejection, edit, escalation or incident rate;
- known vulnerability, prompt injection pattern or approval-bypass issue.

Material oversight changes require testing before release. Tests should confirm reviewer authorization, evidence visibility, decision options, action blocking, version binding, expiry, execution status, logging, escalation and safe failure behavior.

32. Client-Specific Oversight Requirements

Client implementations may require additional oversight rules based on business process, contract, data sensitivity, industry or client governance. Client-specific requirements shall be documented without weakening KOMGIS's minimum controls.

- client reviewer roles and authorized approval groups;
- client-defined approval thresholds or dual-approval requirements;

- source systems or documents considered authoritative;
- communication types that require client approval;
- hours of operation or reviewer availability constraints;
- client escalation contacts and incident coordination;
- retention, export and audit requirements for review records;
- client change-control process for oversight rules.

Where KOMGIS operates the review workflow but the client supplies the human reviewer, responsibilities for reviewer authorization, training, availability, escalation and final action approval shall be clearly allocated.

33. Audit, Management Review and Continual Improvement

Governance review or internal audit should examine a representative sample of systems and completed oversight records. Review should consider:

- whether required systems have a current Human Oversight Plan;
- whether oversight levels match risk and actual action capability;
- whether reviewers are authorized and competent;
- whether the reviewer receives sufficient evidence;
- whether approval gates technically block sensitive actions;
- whether final execution matches the approved version;
- whether material edits, rejection and escalation are traceable;
- whether retention and reviewer data access follow policy;
- whether monitoring detects automation bias, workload and bypass concerns;
- whether incident and corrective-action lessons are reflected in oversight design.

Management review should consider recurring oversight failures, resource and competence needs, changes in AI autonomy, client requirements, high-impact use cases and opportunities to reduce risk through better system design rather than simply adding reviewer workload.

34. Record Ownership, Review and Enforcement

AI Governance / Technical Leadership owns this master record and template set. System owners own system-specific Human Oversight Plans and completed review records for the systems assigned to them.

This document shall be reviewed at least annually and additionally when:

- a new agent or sensitive action capability is introduced;
- a human approval step is added, materially changed or removed;
- a high or critical AI incident involves human oversight;
- review monitoring identifies material automation-bias, workload or bypass concerns;
- new client contractual requirements materially affect approval or review;
- relevant legal, regulatory, security or industry requirements change;
- audit or management review identifies a gap in oversight design or evidence.

Deliberately bypassing a required approval gate, sharing reviewer credentials, approving without required authority or suppressing a good-faith oversight concern may result in access removal, workflow suspension, corrective training, project escalation or other action consistent with KOMGIS's contractual and employment arrangements.

Appendix A. Oversight Level Matrix

Factor	O0 - Normal judgment	O1 - Defined review	O2 - Mandatory approval	O3 - Enhanced / dual approval
Impact	Minor internal inconvenience.	Rework or client dissatisfaction.	Material client or business impact.	Serious legal, financial, safety, privacy or rights impact.
Action	No external or material action.	Read-only or reversible support.	Controlled write or external action.	High-impact, bulk, elevated, irreversible or difficult-to-reverse action.
Data	Public or non-sensitive.	Confidential client data.	Sensitive personal or important client data.	Regulated, highly sensitive or large-scale exposure risk.
Evidence	Normal user context.	Sources / critical facts available.	Defined evidence package required.	Enhanced evidence, warnings and specialist context.
Reviewer	Normal user.	Defined reviewer guidance.	Authorized reviewer.	Authorized specialist and/or second approver.
Timing	During normal use.	Sampling, exception or targeted review.	Before sensitive output/action is released.	Before action, with enhanced escalation and control ownership.
Record	Not normally required.	Reduced review/sampling record.	Complete review record.	Complete review and approval chain; enhanced audit evidence.
Stop / override	Normal user control.	Correction and escalation.	Reject, override and block execution.	Reject, override, pause/disable and incident escalation.

Classification rule: use the highest level indicated by the most consequential reasonably foreseeable scenario unless a documented risk assessment justifies another design. O3 does not mean a system is legally classified as “high-risk”; it is an internal KOMGIS oversight level.

Appendix B. Human Oversight Plan Template

Complete one plan for each production AI system that has a defined human-review or approval requirement. Link the plan to the AI System Inventory and risk assessment.

Field	Value / Required entry
Oversight Plan ID	[HO-PLAN-_____]
AI system name / ID	[System name / inventory ID]
System version / release	[Version]
AI system owner	[Name / role]
Technical owner	[Name / role]
Client / deployment context	[Internal / client / workspace / project]
Intended purpose	[Approved intended purpose]
Internal risk classification	[Low / Limited / High / Critical]
Oversight level	[O0 / O1 / O2 / O3]
Plan effective date	[Date]
Next review date	[Date / trigger based]

B.1 Oversight triggers

Trigger / output / action	Oversight level	Pre-action?	Review frequency / rule
[Describe exact output or action]	[O1/O2/O3]	[Yes/No]	[Every item / sampling / exception / threshold]

B.2 Reviewer roles and authorization

Reviewer role / group	May approve	May override	May stop/disable	Escalation authority
[Role / group]	[Actions / scope]	[Yes/No]	[Yes/No / route]	[Role / contact]

B.3 Evidence package

Evidence item	Required?	How reviewer accesses it
Exact output / action parameters	☐ Yes ☐ No	[UI / record / link]
User request / business instruction	☐ Yes ☐ No	
Source citations / retrieved documents	☐ Yes ☐ No	
Original record / old and new values	☐ Yes ☐ No	
Tool / target / recipient / attachments	☐ Yes ☐ No	

Evidence item	Required?	How reviewer accesses it
Validation and permission status	☐ Yes ☐ No	
Warnings / known limitations	☐ Yes ☐ No	
Prior review or escalation history	☐ Yes ☐ No	

B.4 Decision criteria and critical checks

Critical check	Pass / fail rule	Escalation trigger
[Example: client and case identity]	[Define]	[Define]
[Example: source supports critical fact]	[Define]	[Define]

B.5 Approval-gate behavior

Field	Value / Required entry
Pending-state behavior	[What is blocked and how the item is queued]
Reviewer authorization control	[RBAC / delegated roles / service checks]
Approval binding mechanism	[Version / hash / action ID / immutable payload]
Approval validity / expiry	[Time / state conditions]
Change invalidation rule	[Which changes require new approval]
Execution revalidation	[Checks immediately before action]
Safe-failure behavior	[What happens when approval service / auth / logging fails]
Stop / disable mechanism	[Tool / feature / workflow / credential controls]

B.6 Monitoring and thresholds

Metric	Threshold / expected range	Owner	Response
Approval rate	[Define]	[Owner]	[Investigate / sample / adjust]
Material edit rate			
Rejection rate			
Escalation rate			
Review time			
Bypass attempts	0 expected		
Post-approval incidents			

B.7 Approval of oversight plan

Role	Name / role	Decision	Date / signature or approval reference
------	-------------	----------	--

Role	Name / role	Decision	Date / signature or approval reference
AI system owner		☐ Approve ☐ Reject	
Technical owner		☐ Approve ☐ Reject	
Security / privacy review		☐ Required ☐ Not required ☐ Approve	
Legal / compliance review		☐ Required ☐ Not required ☐ Approve	
Executive / management approval		☐ Required ☐ Not required ☐ Approve	

Appendix C. Sensitive Output / Action Review Record

Use this record for O2 and O3 review-gated items unless the application automatically records equivalent fields in an auditable approval workflow.

Field	Value / Required entry
Review record ID	[HO-REV-____]
AI system / inventory ID	[System / ID]
System / workflow version	[Version]
Model / model provider	[Model / provider where required]
Prompt / workflow version	[Version / release]
Client / workspace / project	[Authorized context]
Case / workflow / task ID	[Identifier]
Requesting user / service	[Identity]
Oversight level	[O2 / O3]
Review trigger	[Sensitive output / action category]
Review opened	[Timestamp]

C.1 Output or proposed action reviewed

Field	Reviewed value
Output / action type	[External message / record update / send / submit / delete / other]
Target / object / recipient	[Exact target]
Tool / integration	[Tool or system]
Material parameters	[Amount / fields / scope / permissions / attachment / destination]
Output / action version or fingerprint	[Version / hash / immutable payload ID]
AI output / action preview reference	[Record / secure link / attachment ID]

C.2 Evidence reviewed

Evidence	Reviewed?	Reference / reviewer note
User request / business instruction	☐	
Source documents / citations	☐	
Original business record / current value	☐	
Validation / schema / business-rule status	☐	
Permission / client context	☐	
Recipient / target / attachment scope	☐	
Warnings / known limitations	☐	
Prior review / escalation history	☐	
Other	☐	

C.3 Reviewer decision

Decision	Select	Reason code(s)	Short note / reference
APP - Approved	☐		
APP-E - Approved after edit	☐		
REJ - Rejected	☐		
OVR - Human override	☐		
ESC - Escalated	☐		
CAN - Cancelled	☐		
INC - Incident suspected	☐		

C.4 Material edits / override details

Change category	Select	Final reviewed value / version reference
No material edit	☐	
Factual correction	☐	
Critical field / value change	☐	
Recipient / target / object change	☐	
Recommendation / conclusion override	☐	
Confidentiality / data-removal change	☐	
Other material edit	☐	

C.5 Reviewer and execution confirmation

Field	Value / Required entry
Reviewer name / identifier	[Name / ID]
Reviewer role	[Authorized role]
Reviewer decision timestamp	[Timestamp]
Second reviewer required?	[Yes / No]
Second review record ID	[If applicable]
Final approval state	[Approved / Rejected / Escalated / Cancelled]
Execution / release timestamp	[Timestamp / N/A]
Execution result	[Completed / failed / blocked / cancelled]
Execution event / audit reference	[Reference]
Incident / nonconformity reference	[If applicable]

Appendix D. Reviewer Checklist

This checklist is a baseline. System-specific critical checks in the Human Oversight Plan take precedence where they are stricter or more specific.

#	Review item	Complete / evidence
1	I confirmed the correct client, workspace, case, person, record or recipient.	☐ _____
2	I understand the intended task and the consequence of approval.	☐ _____
3	I reviewed the exact output or action parameters that will be used or executed.	☐ _____
4	I inspected the required source evidence, citations or original record.	☐ _____
5	Material facts, names, dates, amounts, classifications and targets are correct.	☐ _____
6	I checked for important omissions or misleading lack of context.	☐ _____
7	The item is within the approved intended purpose and business instruction.	☐ _____
8	The information and action are authorized for the active client and user context.	☐ _____
9	I found no unnecessary or unauthorized personal, confidential or cross-client data.	☐ _____
10	Required business rules, validation and process criteria are satisfied.	☐ _____
11	For external communication, recipient, content and attachments are appropriate.	☐ _____
12	I considered whether uncertainty, conflicting evidence or high impact requires escalation.	☐ _____
13	I did not approve solely because the AI output was confident or well written.	☐ _____
14	I found no sign of prompt injection, wrong-client retrieval or unexpected tool behavior.	☐ _____
15	Any material edit or human override is recorded and linked to the final version.	☐ _____
16	My decision code and reason code accurately describe the review outcome.	☐ _____
17	I know how to stop, disable or escalate the workflow if an incident is suspected.	☐ _____

Appendix E. AI Agent Action Approval Record

Field	Value / Required entry
Agent action approval ID	[AG-APP-____]
AI system / agent name	[System / agent]
Agent / workflow version	[Version]
Client / workspace	[Context]
Task / run ID	[Identifier]
Requesting user / service	[Identity]
Tool requested	[Tool name]
Tool scope / credential class	[Read / write / privileged / service account]
Action type	[Send / update / submit / delete / export / other]
Target object / system	[Exact target]
Action parameters / payload reference	[Immutable payload / action reference]
Reversibility	[Reversible / partially reversible / difficult / irreversible]
Oversight level	[O2 / O3]

E.1 Agent-specific approval checks

Check	Status / evidence
The requested tool is approved for this AI system and intended purpose.	☐ Pass ☐ Fail _____
The requesting user and client context are authorized for the target action.	☐ Pass ☐ Fail _____
The agent is using the correct tool scope and least-privilege credential.	☐ Pass ☐ Fail _____
The target, recipient, object or record is correct.	☐ Pass ☐ Fail _____
The exact payload / field changes / attachment scope are visible and correct.	☐ Pass ☐ Fail _____
The action has not been split or rephrased to bypass an approval requirement.	☐ Pass ☐ Fail _____
The consequence and reversibility are understood.	☐ Pass ☐ Fail _____
No suspicious prompt injection or unexpected tool behavior is present.	☐ Pass ☐ Fail _____
Required second approval or threshold approval has been satisfied.	☐ Pass ☐ Fail ☐ N/A _____

E.2 Decision and execution

Field	Value / Required entry
Reviewer	[Name / role]
Decision code	[APP / APP-E / REJ / OVR / ESC / CAN / INC]
Reason code(s)	[Codes]
Material edit / parameter change	[Describe or reference final payload]
Second reviewer	[If applicable]
Approval timestamp	[Timestamp]
Execution status	[Blocked / completed / failed / cancelled]

Field	Value / Required entry
Tool execution event ID	[Reference]
Post-execution verification	[Required? Result / reference]
Incident / nonconformity reference	[If applicable]

Appendix F. RAG / Document-Grounded Review Record

Field	Value / Required entry
RAG review record ID	[RAG-REV-____]
AI system / project	[System / project]
Client / workspace	[Authorized context]
User / task ID	[Identity / task]
Query / request reference	[Reference]
Answer / output version	[Version / fingerprint]
Retriever / workflow version	[Version]
Model / provider	[Model / provider where required]
Oversight level	[O1 / O2 / O3]

F.1 Source and answer review

Review item	Status	Evidence / source reference
Retrieved sources belong to the correct client / workspace.	☐ Pass ☐ Fail	
The reviewer is authorized to access the sources displayed.	☐ Pass ☐ Fail	
Citations identify the supporting document and source location.	☐ Pass ☐ Fail ☐ N/A	
Critical claims are supported by cited or authoritative source material.	☐ Pass ☐ Fail	
The answer does not materially overstate what the sources say.	☐ Pass ☐ Fail	
Conflicting sources or material uncertainty are identified.	☐ Pass ☐ Fail ☐ N/A	
No unauthorized or wrong-client source appears in the answer or context.	☐ Pass ☐ Fail	
No suspicious instruction from retrieved content appears to have changed system permissions or tool behavior.	☐ Pass ☐ Fail	
Material omissions or unsupported claims have been corrected or escalated.	☐ Pass ☐ Fail ☐ N/A	

F.2 Decision

Field	Value / Required entry
Reviewer / role	[Name / role]
Decision code	[APP / APP-E / REJ / OVR / ESC / INC]
Reason code(s)	[Codes]
Material edits	[None / describe / final version reference]
Wrong-client or unauthorized-source concern?	[Yes / No; if yes, stop and escalate]
Incident / nonconformity reference	[If applicable]
Decision timestamp	[Timestamp]

Appendix G. Escalation and Second Review Log

Escalation ID	Source review ID	Reason	Escalated to	Second decision	Date / reference
[ESC-____]	[HO-REV-____]	[Code / short description]	[Role / specialist]	[APP / REJ / OVR / INC]	[Date / record]

Minimum escalation information should identify the original review, reason, receiving authority, resolution, any changed output or action, and whether an incident or nonconformity was opened.

Appendix H. Reviewer Competence and Authorization Record

Field	Value / Required entry	
Reviewer name / identifier	[Name / ID]	
Role / department / client role	[Role]	
Authorized AI system(s)	[Systems / scope]	
Authorized client / workspace scope	[Scope]	
Maximum oversight level	[O1 / O2 / O3]	
Permitted approval categories	[Communication / record update / financial / agent / other]	
Authorization start date	[Date]	
Authorization expiry / review date	[Date]	
Authorization owner	[Name / role]	
Competence requirement	Evidence / completion	Date
System intended purpose, capabilities and limitations	[Training / briefing / assessment]	
Source and evidence verification		
Business process / domain criteria		
AI literacy, hallucination and automation bias		
Prompt injection and suspicious AI behavior awareness		
Privacy, confidentiality and client isolation		
Decision codes, escalation and incident reporting		
Stop / disable / emergency intervention route		
Role-specific or specialist competence		

Authorization decision

Decision	Authorized by	Date / approval reference
☐ Authorized ☐ Authorized with limits ☐ Not authorized ☐ Suspended	[Name / role]	[Date / reference]

Appendix I. Oversight Metrics and Threshold Register

Metric / signal	Calculation / source	Expected range or threshold	Review frequency	Owner	Defined response
Approval rate	[Approved / reviewed]	[Define]	[Weekly / monthly]	[Owner]	[Action]
Material edit rate	[Edited approvals / reviewed]				
Rejection rate	[Rejected / reviewed]				
Escalation rate	[Escalated / reviewed]				
Median review time	[Review decision - opened]				
Very short review percentage	[Below defined duration]				
Queue backlog	[Pending count / age]				
Approval bypass attempts	[Security / workflow events]	0 expected			
Expired / invalid approvals	[Invalidated approvals]				
Post-approval incidents	[Incidents linked to approved items]				
Reviewer disagreement	[Second-review disagreement]				

Threshold breach disposition:

Field	Value / Required entry
Metric / period	[Metric and monitoring period]
Threshold breach date	[Date]
Initial assessment	[What changed / possible cause]
Immediate containment	[If required]
Investigation owner	[Owner]
Corrective action / ticket	[Reference]
Risk / test / oversight plan update	[Reference]
Closure evidence	[Evidence]

Appendix J. Decision and Reason Codes

Decision codes describe the outcome. Reason codes describe why the outcome occurred. Systems may add use-case-specific codes, but the meaning must remain documented and stable enough for monitoring.

Reason code	Meaning	Typical disposition
R01	Incorrect material fact	APP-E / REJ
R02	Unsupported claim / insufficient source	APP-E / REJ / ESC
R03	Material omission or misleading lack of context	APP-E / REJ
R04	Wrong client / workspace / case / person / record	REJ / INC
R05	Unauthorized or excessive data exposure	REJ / INC
R06	Wrong recipient, target or attachment scope	APP-E / REJ
R07	Business-rule or validation failure	REJ / ESC
R08	Sensitive-domain or authority escalation required	ESC
R09	Conflicting evidence / unresolved uncertainty	ESC / REJ
R10	Prompt injection or suspicious content behavior	INC / REJ
R11	Unauthorized tool or permission scope	INC / REJ
R12	Approval mismatch / action changed after review	EXP / REJ
R13	AI recommendation overridden by human judgment	OVR
R14	Task cancelled or no longer required	CAN
R15	Policy, privacy, security or contractual concern	ESC / REJ / INC
R16	Other documented reason	Any; add short note without unnecessary sensitive data

Appendix K. Worked Oversight Examples

The examples illustrate use of the record and do not replace system-specific risk assessment or approval criteria.

K.1 AI-drafted client email

Element	Example
Use case	AI drafts a response to a client about the status of a project and proposes sending the email.
Risk / action	External communication; wrong facts or accidental disclosure could affect client trust.
Oversight level	O2 - mandatory single approval before send.
Evidence	Client/project identifier, user request, relevant project records, exact recipient, email body and attachments.
Reviewer checks	Correct client; facts supported; no cross-client data; no unsupported commitment; recipient and attachments correct.
Decision	APP-E after the reviewer corrects an incorrect delivery date. Reason R01. Final edited version is bound to approval.
Execution	Send tool revalidates recipient and approved message version, then sends and logs the result.

K.2 Enterprise RAG answer with citations

Element	Example
Use case	Employee asks a document intelligence system about an internal procedure.
Risk / action	Read-only document answer; confidential content; errors cause rework but no direct sensitive action.
Oversight level	O1 - defined review and sampling. User can inspect citations and correct or escalate.
Evidence	Inline citations to authorized source documents and source locations.
Reviewer sample	Quality reviewer checks client isolation, citation support, material omissions and unsupported claims.
Monitoring	Unsupported-claim rate, citation correctness, user escalations and wrong-source alerts.
Escalation	Any wrong-client source is R04/R05 and triggers incident review rather than normal correction.

K.3 Agent proposes deleting an important client record

Element	Example
Use case	An AI agent identifies what it believes is a duplicate client record and proposes deletion.
Risk / action	Difficult-to-reverse data deletion with material client impact.
Oversight level	O3 - enhanced approval; deletion remains blocked.
Evidence	Client/workspace, record identifiers, duplicate rationale, side-by-side record comparison, retention status, recovery capability and tool parameters.
First reviewer	Confirms duplicate indicators but escalates because one record contains unique audit history. Decision ESC, reason R09.
Second reviewer	Overrides the deletion proposal and selects a manual merge process. Decision OVR, reason R13.
Control result	Delete tool is never executed. The agent cannot retry with another delete endpoint or split the action to bypass approval.

Reference Documents

- KOMGIS Responsible AI and AI Management Policy - company-wide AI governance framework and human oversight requirements.
- KOMGIS AI Security Policy - approval gates, agent tools, access control, prompt injection and secure development controls.
- KOMGIS Data Privacy Policy - client and personal data handling, retention, deletion and data requests.
- KOMGIS AI Incident Response Plan - containment, investigation, corrective action and reactivation of affected AI features.
- KOMGIS AI System Inventory - system ownership, intended use, models, data, risk and operational status.
- KOMGIS AI Risk Assessment / Impact Assessment - known and foreseeable risks, affected parties, controls and residual risk.
- KOMGIS Model Provider Disclosure / Vendor List - third-party AI and cloud processing records.
- KOMGIS Evaluation and Testing Records - pre-deployment and ongoing quality, security and safety evidence.

Interpretation questions shall be escalated to AI Governance / Technical Leadership and, where legal or regulatory interpretation is required, to qualified legal or compliance review.