

PRACTICAL AI / RESPONSIBILITY AND INTERVENTION

Human Oversight Matrix for AI Workflows

Map where a person monitors, reviews, approves or stops an AI-assisted workflow.

Start with the task, not the tool. Match oversight to autonomy, stakes, data and possible impact if an output is wrong. Use one copy of the working record for each workflow.

USE THIS AS A PRACTICAL OPERATING AID

It is not legal advice, a compliance test or a guarantee that a workflow is safe.

01 Three oversight modes

Monitor	A person samples outputs, checks for drift and records exceptions. Suitable only where errors are recoverable and the workflow does not act on its own.
Human review	A person checks the output before it is used. They can edit, reject or request more evidence before the workflow continues.
Mandatory approval	A named role must approve before a consequential action proceeds. Keep the decision, evidence and authorisation record.

How to use the record

1. Define the task, intended output, data and people affected.
2. Choose the mode and name the accountable owner and overseer.
3. Set checks, approval thresholds, pause, rollback and fallback steps.
4. Record evidence, incidents, changes and the next review.

A responsible person remains accountable even where AI helps draft, recommend or act.

Fill the shaded fields in a PDF reader that supports forms, or print and write. Save a copy, close it and reopen it to check your entries. Keep answers brief; use a separate sheet for more detail. Do not enter passwords or other secrets.

HUMAN OVERSIGHT / FICTIONAL EXAMPLE

Read a worked oversight record

Customer enquiry triage: AI drafts a suggested category and reply for staff review. It does not automatically send a message or make an eligibility decision.

ILLUSTRATIVE EXAMPLE - FICTIONAL ORGANISATION, NOT CLIENT EVIDENCE

Intended use	Draft support only; no automatic sending or eligibility decision.
Data rule	Do not enter personal or sensitive information into a publicly available GenAI tool.
Oversight	Mandatory human review and approval before any reply is sent.
Owner / reviewer	Customer Operations Lead / assigned service adviser.
Review triggers	Incomplete, unusual, sensitive, complaint-related or otherwise uncertain enquiry.
Intervention	Do not send the draft; edit or discard it; use the manual triage process.
Fallback	Standard manual triage template and escalation route.
Evidence	Approved response guidance and the original enquiry record.

The original enquiry stays in its authorised record system. Refer to it when reviewing; an evidence reference is not permission to upload it to the AI tool.

Fictional example, not client evidence

A draft promises an unsupported refund. The adviser stops the draft, checks the original enquiry and follows the manual escalation route. Nothing is sent until required review and approval are complete.

This example shows the kind of control to record. It does not establish that the control is sufficient for every business, sector, decision or legal obligation.

HUMAN OVERSIGHT / WORKING RECORD

Define the work and responsible people

Use short factual entries. Autonomy describes what AI may do; oversight describes what the person must do.

If you cannot describe the intended use or accountable owner, do not move to implementation.

Workflow / recurring task

Intended output

Intended use and stopping point

Foreseeable misuse or misunderstanding

Autonomy: observe, draft, recommend or act?

What can happen without a person?

Stakes and impact if wrong, incomplete or late

Data classification

Public / internal / confidential / personal / sensitive

People affected by the output or decision

Accountable owner

Named overseer who checks, approves or intervenes

Oversight mode

☐ Monitor ☐ Human review ☐ Mandatory approval

Why does this mode suit the autonomy, impact and data?

HUMAN OVERSIGHT / BEFORE THE OUTPUT IS USED

Make the review and approval explicit

Name the check, the evidence and the point at which the workflow must wait for a person.

Pre-use check

What must be checked before the output is used?

Evidence / source

What source, record or test supports the output?

Review trigger

What requires human review or approval?

Accuracy check

How will the output be checked before use?

Approval threshold

Which outputs or actions cannot proceed without approval?

A useful entry says who checks what, against which source, before which action. "A human is involved" does not explain the control.

HUMAN OVERSIGHT / STOP, RECOVER AND RESPOND

Make intervention and recovery usable

The named overseer should be able to follow these steps when a case is wrong, uncertain or unavailable.

Pause / override: who may stop or change the workflow, and how?

Rollback / shutdown: how is the AI step disabled or reversed?

Fallback: what manual or alternative process keeps work moving?

Monitoring record: what is checked, by whom and at what cadence?

Incident / escalation: where are errors, concerns or suspected misuse reported?

Feedback / redress: how can an affected person question, challenge or seek correction where relevant?

Walk through one normal case and one exception. If the overseer cannot find the evidence or use the stop and fallback steps, make those steps clearer before relying on the control.

HUMAN OVERSIGHT / LIFECYCLE AND DECISION LOG

Keep evidence and revisit the workflow

Record responsibilities, reviewer capability and the changes that make the current control unsuitable.

Provider responsibility

What must a provider or contractor explain, support or notify?

Overseer training

Which limitations and failure points must the reviewer understand?

Acceptance evidence

What testing or authorisation supports use?

Review trigger / date

What change or date causes the next review?

Decommission trigger

When should the workflow be retired or replaced?

Decision log: what changed, why, and who authorised it?

Record completed by and date

Approved version or evidence reference

Before relying on the record, check that owner and overseer are named, approval boundaries are explicit, recovery steps are usable, and evidence and the next review can be found.

Keep prior approved versions and the reason for a change where auditability or safety matters. Escalate material privacy, security, legal or sector-specific uncertainty to qualified advice.

Continue with the AI Workflow Starter Checklist