AdriaMont
INSTITUTE

FREE RESOURCE

Responsible AI Feedback Checklist

A human review framework for AI-supported assessment, learning, and development feedback.

Designed for adult professionals and curious learners. Original AdriaMont Institute material, adapted from expert teaching, research, and professional practice.

A practical checklist for reviewing AI-generated feedback before it reaches learners, employees, clients, or stakeholders.

USE THIS WHEN	AI helps draft assessment feedback, summaries, reports, or learning advice
DIFFICULTY	Introductory
FORMAT	Workflow checklist, risk map, stop rules, and review log prompts

Source and privacy note: Designed from AdriaMont's assessment, feedback, and digital learning quality principles. It combines measurement literacy, pedagogical judgment, privacy awareness, and human-in-the-loop review.

The rule of responsible feedback

AI can make feedback faster, clearer, and easier to adapt, but it can also make weak interpretations look polished. The risk is not only factual error. The risk is that feedback becomes too confident, too generic, too personal, or too disconnected from the evidence.

Responsible AI feedback is reviewed as a professional communication. It should be evidence-based, development-oriented, privacy-aware, fair in tone, and clear about uncertainty.

Before using AI

PREPARATION STEP	WHAT TO DEFINE	WHY IT MATTERS
Purpose	Advice, explanation, summary, coaching prompt, report paragraph, or decision support.	Different purposes require different evidence standards.
Evidence boundary	What the model may use and what it must not infer.	Prevents invented claims and hidden assumptions.
Data minimization	Which personal details are truly needed.	Protects privacy and keeps feedback focused.
Audience	Learner, employee, manager, client, team, or public user.	Tone and level must fit the recipient.
Reviewer	Who is accountable for approving, revising, or rejecting the output.	Responsibility remains human.

Five review lenses

LENS	PASS CONDITION	REVISION ACTION
Evidence	Every substantive claim is grounded in available information.	Remove, qualify, or request additional evidence.
Construct	Feedback refers to the intended skill, behavior, or learning outcome.	Rewrite vague or off-construct statements.
Tone	Language is respectful, specific, and development-oriented.	Remove labels, blame, shame, exaggeration, or motivational fluff.
Privacy	No unnecessary personal, sensitive, or identifying detail appears.	Delete, generalize, or anonymize details.
Action	Next steps are realistic, concrete, and proportionate.	Add practice guidance, examples, or a smaller next step.

Stop rules

Do not send feedback if the AI output introduces facts that were not in the evidence.

Do not send feedback if it makes a high-stakes recommendation without qualified human review.

Do not send feedback if uncertainty is hidden or replaced with confident language.

Do not send feedback if sensitive personal information appears without a clear, legitimate need.

Do not send feedback if the advice would be harmful, unrealistic, discriminatory, or outside the competence of the service.

Review log prompts

OUTPUT REVIEWED	What feedback, report paragraph, summary, or recommendation was generated?
EVIDENCE USED	What evidence was the AI allowed to use?
UNSUPPORTED CLAIMS	Which claims were removed, checked, or qualified?
PRIVACY DECISION	What personal details were removed or generalized?
HUMAN REVISION	What did the reviewer change and why?
FINAL DECISION	Approve, revise again, reject, or escalate?

Next step: use the Responsible AI Feedback Review Template for a complete audit trail.