SAGE

Assessment Rubric

Unit: _____ **Task:** _____ **Total marks:** _____

Criterion	Weight	Excellent	Good	Satisfactory	Unsatisfactory
Justification quality & evidence 25–30%	25–30%	Justifications consistently cite specific standards, regulations, or peer-reviewed sources by name. Reasoning is precise and directly addresses the anchor. Modifications are traceable to specific evidence.	Justifications cite relevant sources in most decisions. Reasoning is clear and connected to the task context, with occasional gaps in specificity.	Justifications reference general principles or unit materials but lack specific citations. Reasoning is present but surface-level.	Justifications are absent, generic ("AI was incomplete"), or unrelated to the task anchors. No evidence of evidence-based reasoning.
Gap and error detection 20–25%	20–25%	Identifies AI errors and omissions that require disciplinary knowledge to detect — missing regulatory requirements, context-specific constraints, boundary cases, or exceptions. Rejects AI advice with precise reasoning.	Identifies most significant AI gaps. Some boundary cases or exceptions may be missed. Rejection decisions are generally appropriate.	Identifies surface-level gaps (missing information) but misses deeper disciplinary errors. Acceptance rate is high with limited critical discrimination.	Accepts the majority of AI output without meaningful critique. Gap detection is absent or trivially superficial.
Contextual reasoning 15–20%	15–20%	Demonstrates consistent ability to override AI based on stakeholder context,	Shows contextual awareness in most decisions. Occasionally applies generic	Some contextual adjustments are made but they are inconsistent or not fully justified in	Applies AI output without adjustment to the specific context. No evidence that

		environmental constraints, or discipline-specific requirements the AI could not know. Contextual justifications are precise.	reasoning where context-specific reasoning was needed.	relation to the specific task context.	contextual factors were considered.
Standards mapping **15–20%**	**15–20%**	Accurately maps AI output to named standards, frameworks, or professional codes relevant to the discipline. Identifies specific clauses, controls, or requirements where applicable.	Maps to relevant standards in most decisions. Mapping is accurate but may lack specificity at the clause or control level.	References the existence of relevant standards but does not demonstrate precise alignment. Mapping is approximate.	No reference to standards or frameworks, or references are incorrect and not relevant to the task context.
Metacognitive reflection **10–15%**	**10–15%**	Reflection articulates specific patterns of AI strength and failure observed during the task. Identifies transferable anchoring judgements applicable to professional practice beyond this assessment.	Reflection identifies AI limitations with reasonable specificity. Some connection to professional practice is made.	Reflection describes the process but does not critically evaluate AI performance. Focus is on workflow convenience rather than AI limitations.	Reflection is absent, generic, or focuses on task completion rather than learning about AI capabilities and disciplinary standards.
Output quality **10–15%**	**10–15%**	Final output is professionally presented, complete, and accurate. Could be used in the relevant professional or academic context.	Output is mostly complete and accurate with minor gaps or presentation issues.	Output is adequate but incomplete or contains errors that require correction.	Output is significantly incomplete, inaccurate, or not appropriate for the professional or academic context.

Key principle: 70–85% of marks assess thinking and reasoning (the process). Only 10–15% assess output quality (the product AI can produce).