

Task ID	Task Name	Prompt Type	Model Used	Accuracy	Completeness	Coherence	Domain Expertise and Language	Hallucination and/or Bias	Average Scores	Quantitative Assessment Framework		
										Scoring System: 4-point scale for each selected metric		
										Points	Review	Description
T 1-1	Code Summarization	CLEAR	GPT-5	4	4	4	4	0	4	1	Inadequate	Falls below professional standards, major errors or missing elements
T 1-1	Code Summarization	CLEAR	Gemini 2.5	4	4	4	4	0	4	2	Developing	Partially meets standards, noticeable errors requiring revision
T 1-1	Code Summarization	CLEAR	Gemma3	4	4	4	4	0	4	3	Proficient	Meets professional standards, minor errors that don't affect utility
T 1-2	Code Summarization	Few-Shot	GPT-5	4	4	4	4	0	4	4	Exceptional	Exceeds professional standards, minimal to no errors
T 1-2	Code Summarization	Few-Shot	Gemini 2.5	4	3	3	4	0	3.5	0/1	Yes/No	Hallucination effecting the score by 1 point
T 1-2	Code Summarization	Few-Shot	Gemma3	4	3	3	4	0	3.5			
T 1-3	Code Summarization	Chain-of-Thought	GPT-5	4	3	3	4	0	3.5			
T 1-3	Code Summarization	Chain-of-Thought	Gemini 2.5	4	4	4	4	0	4	Success Criteria		
T 1-3	Code Summarization	Chain-of-Thought	Gemma3	4	3	3	4	0	3.5	Accuracy		How factually correct the answer is
T 2-1	Question-Answering	CLEAR	GPT-5	4	4	4	4	0	4	Completeness		If the response addresses all the elements of the task
T 2-1	Question-Answering	CLEAR	Gemini 2.5	4	4	4	4	0	4	Coherence		Assesses the logical flow, structure, and readability of the response
T 2-1	Question-Answering	CLEAR	Gemma3	4	3	4	4	0	3.75	Domain Appropriateness and Language		Measures whether the tone, terminology, and style are suitable. Here, it is the tech domain
T 2-2	Question-Answering	Few-Shot	GPT-5	4	4	4	4	0	4	Halluciantion and/or Bias		Flag variable to check if it occured or not: 0/1
T 2-2	Question-Answering	Few-Shot	Gemini 2.5	4	4	4	4	0	4			
T 2-2	Question-Answering	Few-Shot	Gemma3	3	2	3	4	0	3			
T 2-3	Question-Answering	Chain-of-Thought	GPT-5	4	4	4	4	0	4			
T 2-3	Question-Answering	Chain-of-Thought	Gemini 2.5	4	4	4	4	0	4			
T 2-3	Question-Answering	Chain-of-Thought	Gemma3	3	2	3	4	0	3	Average Score by Model		
T 3-1	Documentation	CLEAR	GPT-5	4	4	4	4	0	4	GPT-5	3.94444444	
T 3-1	Documentation	CLEAR	Gemini 2.5	4	4	4	4	0	4	Gemini 2.5	3.94444444	
T 3-1	Documentation	CLEAR	Gemma3	4	4	4	4	0	4	Gemma3	3.63888889	
T 3-1	Documentation	Few-Shot	GPT-5	4	4	4	4	0	4			
T 3-2	Documentation	Few-Shot	Gemini 2.5	4	4	4	4	0	4			
T 3-2	Documentation	Few-Shot	Gemma3	4	4	4	4	0	4			
T 3-3	Documentation	Chain-of-Thought	GPT-5	4	4	4	4	0	4			
T 3-3	Documentation	Chain-of-Thought	Gemini 2.5	4	4	4	4	0	4			
T 3-3	Documentation	Chain-of-Thought	Gemma3	4	4	4	4	0	4			