

## Evaluation Questionnaire

<p><b>Coherence    logical</b></p>	<p>response forms a narrative without contradictions. E.g. Is the response well-structured in terms of content?</p>	<p>5 = clearly structured line of reasoning; individual statements are understandable and without contradictions</p> <p>4</p> <p>3 = partly structured line of reasoning; individual statements are confusing or contradicting</p> <p>2</p> <p>1 = unstructured/missing line of reasoning; confusing or contradicting statements</p>
<p><b>Coherence    stylistic</b></p>	<p>Consistent use of visual text formatting E.g. is highlighting or bullet points used consistently?"</p>	<p>5 = consistent use of formatting</p> <p>4</p> <p>3 = some consistent use of formatting</p> <p>2</p> <p>1 = no consistent use of formatting</p>

<p><b>Coverage      broad</b></p>	<p>How well does a response treat a user's contextual information need? Does it cover relevant aspects regarding the given source and query?</p>	<p>5 = all aspects relevant to the user query have been covered 4 3 = some aspects relevant to the user query have been covered 2 1 = no aspects relevant to the user query have been covered</p>
<p><b>Coverage      deep</b></p>	<p>How well does a response treat a user's contextual information need? Does it provide in-depth and highly informative content regarding the given source and query?</p>	<p>5 = depth of information fits the query 4 3 = a little too deep or too shallow for the query 2 1 = too deep or too shallow for the query</p>
<p><b>Consistency   external</b></p>	<p>Consistency between a given source and the produced output</p>	<p>5 = word-by-word consistency / absolutely correct summary 4 3 = no word-by-word consistency / meaning remains the same 2 1 = differences between source and response in terms of meaning/ hallucinations</p>

<b>Consistency language</b>	Consistent language use E.g. Does the writing tone stay consistent?	5 =consistent language use 4 3 = partially consistent language use 2 1 = inconsistent language use
<b>Correctness verifiability</b>	information is verifiable in the source.	5 = information is easily verifiable in the source. 4 3 = some information is verifiable in the source, some is not 2 1 = information is not verifiable in the source
<b>Correctness User intent</b>	information alignes with user's thematic information need (statement fits user intent)	5 = response fits the thematic context of the query 4 3 = response fits the thematic context of the query partially 2 1 = response fails to fits the thematic context of user's needs/query
<b>Correctness language</b>	lexically and grammatically correct language	5 = completely correct language (lexically and grammatically) 4 3 = partially correct language (lexically and grammatically) 2 1 = incorrect language (lexically and grammatically)

<b>Clarity</b>	<b>language</b>	<p>concise, comprehensible, user-accessible and well-spoken language</p> <p>5 = clear language / meeting conversation style 4 3 = unclarity / missing conversation style partially 2 1 = unclear language / missing conversation style</p>
<b>Clarity</b>	<b>saliency</b>	<p>how well key information is placed within the textual response.</p> <p>5 = key information is completely salient 4 3 = key information is partially salient 2 1 = key information is not salient</p>
<b>Cyclicity</b>	<b>content</b>	<p>Does the output contain content-related or literal repetitions?</p> <p>5 = no repetitions (content-related or literal) 4 3 = one repetition (content-related or literal) 2 1 = quite a lot of repetitions (content-related or literal)</p>