



AI Evaluation Rubric

Use the following rubric to evaluate the appropriateness and reliability of AI tools for research purposes. In addition to direct testing, evaluators should review any documentation provided by the vendor about the product. If answers to the following questions cannot be readily determined through testing, additional consultation with the vendor may be required.

Name of tool	
Name of tester	
Date tested	

Contents

Sources	2
Output	3
Functionality	4
User Experience.....	5
Documentation	6
Risks/Benefits.....	7
Recommendation	7

Sources	How would you assess the scope, quality, and currency of the sources used?	
	Breadth: What discipline(s) does this tool support? Within the disciplines it is designed to support, are there any notable gaps in coverage?	
	Quality: What types of sources are included (primary vs secondary, peer-reviewed vs popular, subscription vs open access, etc.)? Are these sources appropriate for the discipline(s) and use cases the tool supports?	
	Depth: Does the tool include the full text of these sources, only limited excerpts (such as metadata and/or abstracts), or some combination?	
	Currency: What is the approximate date range of the sources used? Is this level of currency appropriate for the discipline(s) and use cases the tool supports?	

Output	How would you assess the quality, accuracy, and relevance of the tool's responses?	
	Accuracy: Is the information provided by the tool accurate or credible? Does the tool accurately summarize the findings, methodologies, and significance of the retrieved sources?	
	Depth/Clarity: Are the responses provided by the tool clear, concise, and easy to understand? Do they include sufficient detail to be useful to the user?	
	Relevance: Is the information and sources provided relevant to the user's research topic or question? Are they appropriate for the intended audience (i.e. not too elementary or advanced)?	
	Attribution: Does the tool fully and accurately cite its sources? Does it include inline citations attributing specific parts of its response to the appropriate source?	

Functionality	How would you assess the available features and functionality of the tool?	
	Refinement: Can users clarify or refine a query by asking follow-up questions or applying search filters (e.g., date, author, publication, etc.)?	
	Personalization: Can the tool be tailored to the user's specific needs, such as academic level or discipline?	
	Save/export: Can users easily save or export the response and sources provided by the tool?	
	Share: Can users generate a shareable link to a conversation to share it with others?	
	History: Can users access their chat history and past conversations?	

	Feedback: Are users able to provide feedback regarding the quality of responses? Is this feedback used by the vendor to improve performance?	
	Recommendations: Does the tool offer additional source or query suggestions for further exploration?	
	Out-of-scope queries: How does the tool handle questions that are outside the scope of its training data?	
User Experience	How would you assess the user experience of this tool?	
	Ease of use: Is the tool intuitive and easy to use, particularly for users with varying levels of technical skill?	
	Performance: How quickly does the tool retrieve and display results?	

	Error handling: Does the tool provide meaningful feedback when errors occur, such as when a query yields no results?	
	Accessibility: Is this tool accessible to users with disabilities? Is the tool accessible on desktop and mobile devices?	
	Privacy: Does the tool appropriately safeguard user data and comply with applicable privacy policies? Is any user data used in the training of the tool or shared with third parties? If so, can users opt to exclude their data from being used or shared?	
Documentation	How would you assess the documentation explaining the tool and its use?	
	Completeness: Is there sufficient documentation that clearly explains what this tool is, how it works, and the data sources used to generate responses?	
	Transparency: Does the vendor provide relevant information about the limitations of generative AI and their tool (e.g., bias, misinformation, intellectual property, etc.)?	

	Visibility: Is this documentation easily discoverable and accessible to users at the point of need?	
Risks/Benefits	What risks and/or benefits does this tool offer to the library and its users?	
	Goals: What need or problem is this product intended to solve? How well does it solve it?	
	Benefits: What (if any) benefits does this tool offer to the library and its users?	
	Risks: What (if any) risks does this tool pose to the library and its users?	
Recommendation	Do you recommend that the library acquire or use this tool?	
	Recommendation: Do you recommend that the library acquire or use this tool?	

	Improvements: What improvements do you recommend the vendor consider as they continue to develop this tool?	
--	--	--