

Evaluating generative AI-powered resources for higher education

Parts of a generative AI application

Large Language Model (LLM) is the underlying engine for the chatbot or research assistant. It's a statistical model that predicts the most likely string of words related to the input (also called a prompt). Examples include: GPT-4, Claude, and Gemini.

Retrieval-Augmented Generation (RAG) sits on top of the LLM and acts as a guardrail for content output. It allows users to choose datasets and will often reduce hallucinations. The RAG can also govern how prompt results are displayed: prose, search results, graph, etc.

Open Web AI Resources

Are the responses **transparent**? You should be able to go back to the original document and find where the application got its information. For example: in-line citations, highlighted sections of the original document, etc.

Is the application **efficient**, or does it make more work? A good generative AI application will augment your research workflows, not waste your time with wrong answers and hallucinations. The tool should also match existing and established research workflows, not reinvent them.

Does the tool eliminate too much of the **critical thinking** process? It's important that researchers retain and/or enhance their critical thinking skills as they work through their research.

Are your sessions **private**? Always check the privacy policy for the tools you are engaging with. And never enter personal data into a generative AI application.

Subscription AI resources

Who is the vendor?

Both legacy academic publishers and education technology companies are producing quality generative AI tools for scholarly workflows. Make sure the organization you are doing business with understands the academic honesty and data privacy values of your institution.

Research assistant vs. Chatbot

Some of the generative AI applications best suited for scholarly workflows are research assistants because they do not produce text that may undermine academic integrity policies. These tools are designed to help scholars think cross disciplinary and develop new ways to think about their research question.

Choice's rubric for evaluating AI tools

- What kind of student population does this application serve? Community college, four-year, masters/PhD?
- How does this application fit into an existing library collection?
- How does the application measure up against the developer's or publisher's claims of productivity and performance?

**Learn more about Choice at
Choice360.org**