UNIVERSITY OF CALGARY
Student Success Centre

Evaluating AI Outputs

Our environment is saturated with information and, with the ability of AI tools to produce human-like content within seconds, it can be increasingly difficult to know whether the content we read is reliable. This handout is designed to provide a brief overview of how you can evaluate AI outputs to engage critically with information and claims in AI-generated content.

Additionally, the principles outlined in this document should also help you evaluate content beyond AI outputs, ultimately helping you differentiate between helpful and harmful information.

Asking AI "Are You Sure?"

The simplest way to test the validity of AI generated content is to challenge it. If you are talking to an AI chatbot, challenge its outputs to double check validity. When an AI chatbot gives you a response, simply asking "Are you sure?" is a great first step towards testing the validity of the content.

One effective method for critically evaluating AI outputs is the "R-U-SURE" method. Using this acronym, we can break the process of critical engagement down into 6 steps:

Reflect

Take a moment to reflect on both your input and the AI's output. Allowing time for reflection means you are less likely to base your reaction on initial emotions or assumptions. Give yourself some time to question whether the information in front of you is logical and coherent.

Unpack

Consider how the AI tool responding to your input. Is it blindly supporting your statements? Is it being overly agreeable? Is it providing narrow or diverse perspectives? Why might it be doing that? Is there something you can do on your end to address potential concerns?

Search

AI reproduces claims without fact-checking. If citations are not already provided in the tool's initial response, ask the AI tool to cite its claims, then verify those citations. AI is designed to be convincing and agreeable. This generally means it can present you with plausible but inaccurate information and fabricate citations. Even when an AI lists real sources, it is still a good idea to read the sources for yourself. Simply checking that a source cited by the AI exists does not show whether the AI used it correctly or even consulted the source at all. A common issue that can arise when AI is asked to provide citations, is the creation of post-hoc citations.

Post-hoc citations are sources that an AI tool adds after it has already generated an answer. The AI does not start by checking those sources. Instead, it first produces a response based on patterns in its training, then looks for sources that seem to support the response. It is the user's responsibility to verify claims by investigating citations and checking for accuracy.

Understand

Keep in mind that AI can make mistakes. Read your sources carefully so that you develop your own understanding without relying on AI tools for summaries. Instead, use AI to reinforce your understanding of the topic by interrogating points and comparing diverse perspectives.

Once you have done this, look at the AI output again and identify what might be incorrect, misleading, or missing. Compare the output with your sources. Go back to your input and identify how you can add or remove context to tailor a better output.

Revise

Often when chatting with an AI tool, the first few outputs can be overly broad. You can continue adjust your inputs and approach until you arrive at a relevant output. Additionally, you should think about other tools to help you throughout. For instance, you may also browse databases and search engines, or you may approach librarians and subject matter experts for a more robust process. Avoiding AI as your sole research tool strengthens both your validation process and your overall research approach.

Evaluate

This step emphasizes your personal responsibility to make judgments yourself. Instead of asking AI which sources are reliable, you can use AI to help you identify key points based on popularity and strength of the evidence. The final assessment is yours to make, and you remain fully accountable for the conclusions you draw.

Asking Yourself "Am I Sure?"

When using AI tools, you should also ask yourself, "Am I sure?" Although this handout focuses on evaluating outputs, you should still question when and how you are using AI tools. Here are some good questions to ask yourself about your general AI use:

- Why am I using AI here?
- How am I using AI? What for?
- What viewpoints might I be missing?

When you use AI to help you complete coursework, you should also ask yourself academic questions, such as:

- Do I have explicit permission from my instructor or supervisor to use AI?
- Do I fully understand my instructor's expectations regarding when and how I can use AI to help complete assignments?

- How will I disclose my AI usage?

Lastly, before you upload documents or provide information to the AI, stop to ask yourself:

- Am I allowed to upload this document to the AI tool, or will doing so constitute copyright infringement?
- How is the data I input being used for data collection and/or model training?

The SIFT Method

The SIFT method (Caulfield, 2019) is a framework designed to help people filter out weaker, unreliable sources. Caulfield outlines four steps to maintain a strong research process that emphasizes accuracy and reliability:

Stop

The first move is to pause and question both the claim and the place in which you find the claim. Avoid sharing or believing any information without first pausing to reflect.

Investigate the source

Every source of information you encounter has a purpose, whether it's to educate, inform, or convince you. For example, this handout is designed to help you use AI effectively. Likewise, all information you encounter—whether from AI, in class, on textbooks, online, on social media, on TV, on the news, or from friends—will have a purpose worth investigating to help you make sense of its significance. Therefore, the purpose of any given source is important to consider when you evaluate for certain biases or conflicts of interest.

Find better coverage

The digital age means anybody can reproduce information, so it can be fruitful to look for other sources that cover the same topic. Look for trusted sources and compare different perspectives if possible.

Trace claims, quotes, and media back to the original context

The original source (i.e. the original study, report, proposal, interview, etc.) of a piece of information or claim will provide you with the necessary context to finalize your evaluation of that material. Refer to the original source to determine whether its claims were summarized or reproduced accurately. You should be wary of claims that do not seem to have an official "origin."

SIFT through AI outputs

Applying the SIFT method to AI outputs requires additional steps when verifying sources. AI can inaccurately cite sources or fabricate citations that seem real. You will have to analyze sources by answering the following questions:

- Is AI citing real sources?
- What kind of source is it citing? Is the source academic?
- Is AI actually consulting that source? How do you know?
- Does the AI output faithfully represent the original source?

Validating Sources

Whether a source of information is written by a person or AI-generated, you should always evaluate sources for accuracy and reliability.

Both the R-U-SURE method and the SIFT method can help you interrogate information for validity while developing critical reading and thinking skills. But note that different tools will require different processes of verification.

For instance, tools like ChatGPT and Microsoft Copilot are trained on public internet content with varying levels of credibility. The validation process may involve ensuring that the output carries the same meaning as the source origin. When possible, you should also trace the original source and question its intention, thinking about who wrote the information and why.

On the other hand, tools like scite.ai and Scopus AI limit their training data to scholarly sources like academic journals and peer-reviewed content. The validation process when using these tools may involve reading and understanding the literature cited to develop a critical understanding of the content while avoiding GenAI summaries to prevent "hallucinations."

Additionally, when evaluating AI's citations, keep in mind:

- AI tools normally only access open-source content; they may provide citations they are not actually deriving information from.
- If you **did not** include context in your prompt, the AI tool is forced to rely only on its training data, which means it may generate a generic response based on a wide array of sources.
- If you included context in your prompt, make sure the AI tool is not simply repeating the information you provided.

Mitigating Bias

Bias in AI can manifest in various ways, from showing only one popular viewpoint and ignoring the rest to outright perpetuating stereotypes. Moreover, bias can occur at any stage throughout the AI process, including during data collection.

The way you use AI can also generate biased responses. AI tools are usually designed to meet your needs, and to do so, they must be agreeable. These tools thus tend to agree with you, even if your own inputs are misleading or incorrect.

Again, the R-U-SURE method can help mitigate bias in AI by encouraging critical reflections, especially regarding perspectives that might be overlooked. Ask the AI tool for diverse perspectives and counterarguments, then you can critically evaluate all those outputs.

The Transparency Issue

Just as failing to disclose your AI usage can prove problematic, companies' lack of transparency about how their AI tools work can lead to problems such as unchecked biases.

Unlike traditional algorithms, AI—a machine learning technology—learns an algorithm based on its own calculations and pattern recognitions. This means that, though companies can be transparent about their "learning algorithm," the "algorithm learned" is not always accessible (Zerilli et al., 2021). As a result, it is difficult to pinpoint **how** or **why** an AI tool would generate a given output. You should approach every output with a healthy amount of skepticism.

Conclusion

The world is saturated with information. A good portion of the content we see online or hear in person can be inaccurate. The most reasonable way to assess any piece of information is to give it more thought. Any information can be "made up" and presented convincingly. Consequently, even citations can be inaccurate.

When facing new information, you may benefit from simply taking a moment to think about the logic, reasoning, and feasibility of what you are being presented. This short moment of reflection helps you develop critical thinking and personal judgment skills.

References

Caulfield, M. (2019, June 19). SIFT (The four moves). *Hapgood*. https://hapgood.us/2019/06/19/sift-the-four-moves/

Zerilli, J., Danaher, J., Maclaurin, J., Gavaghan, C., Knott, A., Liddicoat, J., and Noorman, M. (2021). *A citizen's guide to artificial intelligence*. MIT Press.