

Don't Trust AI to Cite Its Sources

By Anna Mills and Maha Bali

Teachers will usually ask you to cite sources, right? They will explain that you should cite not just when you are quoting but also when you are referencing an idea you got from someone else. The hope here is that in academic practice we give some sense of where an idea is coming from. We think of academic work as a big ongoing conversation between people agreeing and disagreeing and building on each other's points as they try to figure things out. It helps to keep track of who is saying what. It seems only fair to give credit to the person who developed each main point.

For some concepts in the humanities and social sciences, it is also important to know the positionality of the person or people who developed them. For example, if we are studying nonviolent resistance, we need to understand the people and context in which ideas about it were developed. Mahatma Gandhi, Martin Luther King Junior, and Henry David Thoreau wrote extensively about it, and their ideas were shaped by very different experiences, times, and places.



Mahatma Gandhi was instrumental in shaping current understandings of nonviolent resistance in 1931, public domain image courtesy of [Wikimedia](#)



Martin Luther King Junior was also instrumental in shaping current understandings of nonviolent resistance. Image by [WikiImages](#) from [Pixabay](#)

Knowing something about how these leaders were positioned in terms of race, gender, class, education, and politics can help us understand their ideas and think about how we want to respond from our own positionality. However, if you ask a chatbot about “nonviolent resistance,” it may give you an answer without reference to any person or context at all (chatbot responses are variable). Here is an [example with ChatGPT](#):

what is nonviolent resistance? Answer briefly



Nonviolent resistance is a method of protest and civil disobedience that uses peaceful tactics, such as marches, sit-ins, and boycotts, to oppose government policies or societal issues without using physical force. This strategy aims to bring about social or political change through moral or economic pressure, often appealing to the conscience of the public and decision-makers.

Screenshot from a [ChatGPT4o temporary session, July 2024](#)

In general, chatbots don't make it easy to figure out what sources have influenced their answers. As [Iris Van Rooij puts it](#), “LLMs, by design, produce texts based on ideas generated by others without the user knowing what the exact sources were.”

The systems don't “know” what influenced their answer. [Dominik Lukeš refers to this limitation](#) as a lack of “introspection.” Once a system is trained, it consists of a very complicated formula, a big set of numbers that its supposed to multiply by other numbers. The chatbot puts our question or request into the formula and gets a result. It has no way to look backward to see which human writings made its formula have certain numbers.

Chatbots often don't cite when they should

Why does it matter if chatbots can't tell us where an idea came from? In June 2023, Maha Bali pointed this out to Anna Mills in a conversation on X. She [wrote](#), “One of the things I'm stuck on right now is that a lot of the AI-generated text paraphrases work of scholars we would *normally* cite as paraphrased. This stuff now goes unacknowledged and it's not OK...E.g. I asked it about characteristics of White Supremacy Culture. It gave me the list that Tema Okun and others use, but did not cite her.”

A year later, I (Anna) tested and saw the same pattern. I asked three chatbots—ChatGPT, Claude, and Gemini—about “Intentionally Equitable Hospitality,” a concept developed by Maha Bali and other co-directors of the grassroots group Virtually Connecting.

- ChatGPT generated long, inaccurate descriptions of the concept with no reference to either the organization or the people who developed it. This was true of all ten times I tested (see [this sample ChatGPT transcript](#)).
- Google's Gemini described the concept vaguely, making up principles that are not in any published writings about this exact phrase, “Intentionally Equitable Hospitality.” I tested it ten times and not once did it mention an organization or person. (See [this sample Gemini transcript](#)). Gemini did feature a Google “G” logo which offered to check the results when I hovered over. The check led it to highlight several passages of Gemini's output and link them to websites, including Maha Bali's blog.
- Claude 3.5's ten responses all referred to the hospitality industry, a clear error since the exact phrase “Intentionally Equitable Hospitality” was developed to refer to education (See [this sample Claude transcript](#)). They also failed to mention any organizations or people involved in developing the concept.

Citing chatbots according to the rules gives them credit for human authors' work

Labeling chatbot output as such is important. And the major academic organizations are working hard to come up with guidelines for this. But it's tricky. If we follow the current guidelines, we'll make no mention of the humans who developed the idea the chatbot is summarizing.

Take any of the examples above where we asked a chatbot about Intentionally Equitable Hospitality. Let's say a person does that and then wants to quote the output and follow the rules to cite ChatGPT in APA, MLA, or Chicago style. All of these would make it look like ChatGPT is the source of the IEH concept. For more on this, see [If I use AI, how should I acknowledge or cite it?](#)

Don't ask a chatbot what its sources are

There is a lot of confusion about this among teachers and the general public because if you ask a chatbot what its source is, it will often give an answer that sounds plausible. It still doesn't “know.” It is using its word prediction abilities to “guess” which source influenced its output. It doesn't really have access to its own internal workings. The source it mentions may not exist or it may not be the one that really shaped the response. For example, in one case, ChatGPT said it got information about Intentionally Equitable Hospitality from [Kimberly Crenshaw](#), who is known for the concept of Intersectionality and has not written about IEH.

Chatbots also respond unreliably when you ask which person developed a concept.

- When I [asked ChatGPT the follow up question](#) about IEH: “Is this a particular person's concept?” it gave credit to Dr. Tia Brown McNair, a prominent expert on Diversity, Equity, and Inclusion who is not in fact associated with the phrase “Intentionally Equitable Hospitality.”
- When I [asked the same follow-up question of Claude 3.5](#), it gave credit to Ashtin Berry, a hospitality industry leader whose name is not associated with the phrase “Intentionally Equitable Hospitality.”
- Google's Gemini did better. It [responded to the question](#) once by correctly referring to Virtually Connecting and its co-directors, including Maha Bali and Autumn Caines. On another trial, it [referred only to Maha Bali](#). On another, it [referred to Maha Bali, Mia Zamora, and the organization Equity Unbound](#), which is correct because both authors and the organization have developed this concept further beyond the original by Virtually Connecting. None of the trials trace the idea from its original formulation to its current formulation.

Even chatbots that can search often don't mention their sources

As we will see in The pros and cons of AI for research assistance (coming soon), many current chatbot/search hybrid systems can search the internet in real time to inform their response to a question. They link to documents, making it seem like they are allowing us to trace the source of their information. But we the users don't know if those links are really the sources they used or

not. And it isn't always clear which of the links informed which part of their response. Besides, many people will not click on the links; they will focus on the AI response itself.

In May 2024, Google [rolled out "AI overviews"](#) at the top of many search results pages. When I try the same query about Intentionally Equitable Hospitality (IEH) in Google search, I get an answer that doesn't mention a source, though links to sources are right below it.



Screenshot of a Google search AI Overview in response to the query "What is Intentionally Equitable Hospitality?" on July 19, 2024. Four repetitions produced similar results.

The Google result above plagiarizes most of a sentence from the first source linked to, Bali and Zamora's "[Intentionally Equitable Hospitality as Critical Instructional Design](#)." Google explains "IEH is iterative and involves planning, design, and facilitation in the moment" without quoting. Bali and Zamora had written, "IEH is iterative design, planning, and facilitation in the moment."

Now let's try with the popular chat-search hybrid and Google alternative Perplexity.ai. In [one instance](#), Perplexity mentions Virtually Connecting as a source for IEH and linked to an article by Maha Bali and Mia Zamora without mentioning their names. (How many will follow the link?) On the [second](#), [third](#), [fourth](#), and [fifth](#) trials with the free version of Perplexity, it [gave links](#) but mentioned no source at all in its overviews. I repeated this five more times using my quota of free "Pro" searches. Only once did it mention [Maha Bali and Mia Zamora](#) by name in its autogenerated answer. So even though this chatbot was referencing a real source by the scholars who defined the concept, in five out of ten trials it included no equivalent at all to an in-text citation (see [all of Perplexity's responses](#)).

So how do we respect sources when using AI?

All this means that when a chatbot "says" something, we should wonder not just whether it is accurate but also whom the chatbot is parroting. Of course, many concepts have been shaped by so many contributors and have become so widespread that they do not need to be cited. For example, if we state, "Many Mexicans have both indigenous and European ancestry," we do not need to cite it.

But if the point is less well known or more controversial, it is left to us to do due diligence and see if there is a particular person or group of persons who came up with the ideas AI serves us. Internet and scholarly database searches on key concepts may lead us to the human thinkers responsible.

Should it really be on users to try to reconstruct where a chatbot may have gotten its information? Is there any way the software itself could help us? If so, how could we encourage the companies or the government to make that happen? See "Does using AI do harm? If so, should we stop using it?"* for more on intellectual property and how we might participate in shaping AI in this regard.

Further reading

- [Acknowledging and Citing Generative AI in Academic Work](#)
- [Find the Conversation That Interests You](#) from the chapter on research in *How Arguments Work*
- [ChatGPT is bad at following copyright law, researchers say](#) by Britney Nguyen, *Quartz*
- [The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work](#), By Michael M. Grynbaum and Ryan Mac, *The New York Times*, December 23, 2023
- [A new bill wants to reveal what's really inside AI training data](#), *The Verge*, April 10, 2024

References

- Understanding AI's limitations and dos and don'ts" by Lukeš, Dominik, included in [Integrating AI into Academic Practice: Guide to Reflective Exploration](#)
- [Against Automated Plagiarism](#) by Iris Van Rooij
- [Intentionally Equitable Hospitality as Critical Instructional Design](#) by Maha Bali and Mia Zamora in *Designing for Care*
- Jackson, K. F., Goodkind, S., Diaz, M., Karandikar, S., Beltrán, R., Kim, M. E., Zelnick, J. R., Gibson, M. F., Mountz, S., Miranda Samuels, G. E., & Harrell, S. (2024). Positionality in Critical Feminist Scholarship: Situating Social Locations and Power Within Knowledge Production. *Affilia*, 39(1), 5-11. <https://doi.org/10.1177/08861099231219848>

[Don't Trust AI to Cite Its Sources](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.