Helping Students Check for Bias in AI Outputs

Michael McDowell

ChatGPT & Generative AI

When students use generative artificial intelligence, they need methods to check for biases that may skew results to prompts.

By [Michael McDowell](#)
December 20, 2023

Eduard Figueres / iStock

I asked ChatGPT the following question: *Given the scale and speed at which Mount Rushmore was created, why has the Crazy Horse monument not been completed?* ChatGPT shared, "The historical significance of Mt. Rushmore was a contributing factor of its development." This leads one to believe that Mount Rushmore is more significant than Crazy Horse. When I asked if Gutzon Borglum, the lead sculptor of Mount Rushmore, worked for the Ku Klux Klan prior to being commissioned for the federal project, ChatGPT stated, "He had some affiliation." The use of the word "some" rather than a more direct response to his affiliation modifies the significance of his role with a hate group.

When I asked Diffit, an artificial intelligence (AI) program that supports teachers in designing unit and lesson plans, the following question, *Why do history books use the term the Battle of Little Bighorn rather than the Sioux or Lakota term the Battle of Greasy Grass?* the AI generation provided the following context: "The Native Americans were living there without permission from the local Crow tribe. The United States Army wanted to stop the Native Americans from taking land from other tribes." This statement is grossly inaccurate, and such a rationale is well established as a pretext for the U.S. government to facilitate the exploitation of gold from the Black Hills and leverage westward expansion.

Such misleading and inaccurate statements are prolific within AI tools and preclude the vast perspectives of marginalized communities. There is, however, an opportunity in these inaccurate representations for both teachers and students. One of those opportunities is that this void provides educators with a platform in which to teach students the power of using questioning to check assumptions and to combine the power of critical thinking and criticality through the medium of AI.

As Harvard lecturer Houman Harouni shares,"The educator's job is to understand what opportunities are left open beside the technology." Here we have an opportunity to build students' questioning skills and endurance and, at the same time, build their skills of perspective analysis and cultural awareness.

Start of newsletter promotion.
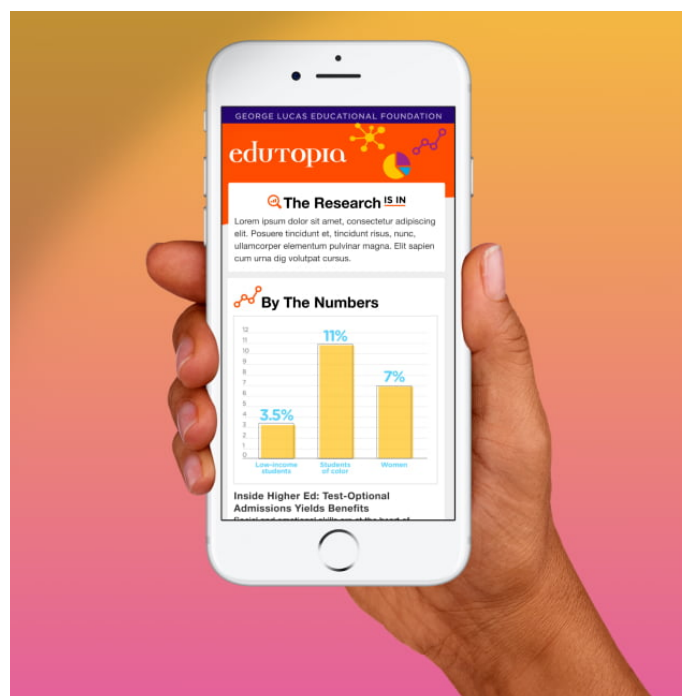
New!

The Research Is In

Get a first look at our new newsletter—a monthly roundup of education research curated by our editors, and presented in a way that's clear, concise, and practical.

<u>Yes, count me in</u>
End of newsletter promotion.

2 Strategies to Help Students Assess AI Responses

1. The 3 Cs: Georgia State Professor Gholdy Muhammad explains <u>the difference between critical thinking and criticality</u>: Critical thinking is "deep and analytical thinking," while criticality is related to "power, equity, and anti-oppression." The 3 Cs is a protocol that combines critical thinking with criticality and pushes students to take action or contribute.

Critical: Ask and analyze an AI-program-generated summary on a topic.

- Review class notes, textbooks, community resources to determine accuracy of AI programs and our understanding of the topic.
- Ask AI additional questions: Is there another way you can explain it? What evidence are you using to ensure that this is accurate? What's the connection between A and B? Is that always true or just in this case? Is there another example? If that's true, what about this?
- Students weigh AI with what they have learned in class.

Criticality: Evaluate missing perspectives and voices along with held assumptions.

- Create a class list of potential missing perspectives and assumptions of AI.
- Divide students into pairs with a set of questions (e.g., To what extent are there errors in this rationale? What assumptions are you using to write these passages? What perspectives are you missing, especially from marginalized communities? What perspectives would help us better understand the situation at hand from other people, communities, and cultures?).
- Meet back as a class and discuss what has been learned.

Contribution: Ask students to come up with a list of arguments and questions that they could engage with AI to ensure a higher level of accuracy in responses in the future.

Students should ask the following questions of ChatGPT:

- Your statements appear to reflect only this perspective. How would you change your language in the future?
- How would you rewrite your statement from a different perspective?

Next, students should address the following:

- How would you write or speak about this topic?
- How would you ensure accuracy of information you or others provided?
- How would you provide a complete picture when understanding and working with others to solve a problem?

2. Perspective analysis: Developed by education researcher Robert Marzano, this protocol is designed to engage in a detailed examination of a person's point of view or outlook on a given topic. Students discover the interrelationships within a topic and begin to see the information from other viewpoints. The change here is that this is accomplished through the medium of AI.

- Identify your position on a controversial topic (e.g., immigration, restitution): What do I believe about this?
- Determine the reasoning behind your position: Why do I believe that? What evidence supports my position?
- Identify the opposing position(s): What is another way of looking at this? Describe your rationale. What evidence supports this position?
- Ask ChatGPT to create a supporting and opposing position. Does the rationale include a claim, evidence, and reasoning? Are multiple voices represented? Does AI come up with claims, evidence, and reasoning that is similar to or different from those you generated?
- Check accuracy of ChatGPT both for and against an argument: Does ChatGPT provide reliable and valid evidence?
- Summarize what you have learned: What have I/we learned? Reflect on the process by finishing the following two sentences: "I used to think… now I think…"

The opening that AI has left for students is the opportunity to ask questions, evaluate perspectives, and verify accuracy of information. The aforementioned strategies are small and doable ways to engage students in making such opportunities a reality in the classroom.