

Who Is Missing from AI's answers—and Why?

Imagine trying to piece together the history of a women's community from what was left behind. A shoebox of dance flyers. A photograph with only a few names written on the back. The minutes from one committee meeting. You ask around and hear about the woman who opened her home, the one who raised the money, the one who quietly brought newcomers into the circle. Much of what held the community together was never written down.

I've been thinking about that as we turn to AI for answers. It can draw on an enormous amount of material, but who made that material? OpenAI says its models are developed using public information, material obtained through partnerships, and information provided by users, trainers and researchers. Each of those sources has already passed through human hands. Someone had the time and means to publish. Someone decided what was worth preserving, translating or making available.

For a long time, the people most able to leave a public record were also the people with the most power. Their speeches were printed. Their work was credited. Their account of an event became the account. Others lived no less fully, but left fewer traces in the places an AI might learn from.

Some women's work was never lost; their credit was. Scientific discoveries were attributed to male colleagues. Paintings were sold as a husband's work, and books published under his name. The experiment, the canvas, the manuscript survived. The record was not silent. It named the wrong person.

Even when a woman's name survives, she may appear in a history as somebody's wife when she was also the person who organized the work. A community may be described by an outsider who never understood what its members valued.

This is why the problem goes deeper than counting who appears in an AI's sources. We must also ask who described them, whose words were treated as credible, and what kinds of work were considered worth naming. The National Institute of Standards and Technology points out that AI bias can come from institutions and social practices as well as from the data itself.

Picture a meeting. A woman offers an idea and it passes without comment. Later a man repeats it, and now people listen. The minutes record the proposal and the decision; they may give no hint of who spoke first. If an AI is asked whether those minutes prove she was ignored, "the evidence is inconclusive" may be a careful answer. But it is careful only about the minutes.

An answer that notices the room around the record would ask who wrote those minutes, whose contributions they usually capture, and whether this happens again. It would not declare a pattern proven by a single account. Nor would it treat a woman's inability to produce a perfect paper trail as evidence that nothing happened. Speaking up might cost her something. Power can shape the event, the telling of it, and what is left for the next person to find.

Other silences deserve care. Speaking openly was unsafe for some people, and some knowledge belongs to a family or community that never chose to publish it. We cannot repair the record by simply taking every private story and feeding it to a machine. People should have a say in whether and how their knowledge is shared.

AI can connect thousands of sources in ways no single person could hold in mind. Yet if those sources share a blind spot, their sheer number can make an old silence sound like consensus. I want an intelligence capable of saying, This is what the available sources tell me. Here is where they fall quiet. Here is whose account I have not found. It should not invent the missing voices. It should make room for us to seek them.

We can help by preserving histories with permission, by asking who has been overlooked, and by involving the people whose lives an AI is meant to understand in decisions about how it is built and tested. One of our responsibilities is to teach it to examine the record itself before calling its silences neutral.

Before accepting an answer about humanity, perhaps we should ask one more question: Who is missing—and why?

In the interest of brevity, I have concentrated on women's voices, but it applies to all voices that do not have the power or resources to be heard.

Kim