JUST TECH

Anti-Transgender Disinformation in the Age of Algorithmic Search Summaries

By: Anna Beers

PLATFORMS & INFRASTRUCTURE

Beers, Anna. "Anti-Transgender Disinformation in the Age of Algorithmic Search Summaries." Just Tech. Social Science Research Council. May 7, 2025. DOI: 10.35650/JT.3083.d.2025.

Say you were a transgender person, or the friend or family of a transgender person, who wanted to learn more about the process of medical transition. Like many people with a question, you might turn to Google Search, entering a query like "long-term effects of transitioning," hoping to find some informative websites. Because of a new Google Search feature, you would find a short AI-generated text directly answering your question, placed before any actual search results. As of writing, the text would seem quite informative, including linked citations to peer-reviewed literature, government medical agencies, university websites, and news outlets. However, it also includes material sourced to the "American College of Pediatricians," an anodyne-sounding organization that is one of the US's most radical purveyors of anti-LGBTQ+ disinformation. How does a group that promotes conversion therapy for gay and transgender children, opposes parenthood for gay adults, and writes skeptically of Covid-19 vaccines get cited as a trusted source in AI-generated search summaries, and is mass-distributed to the open web?

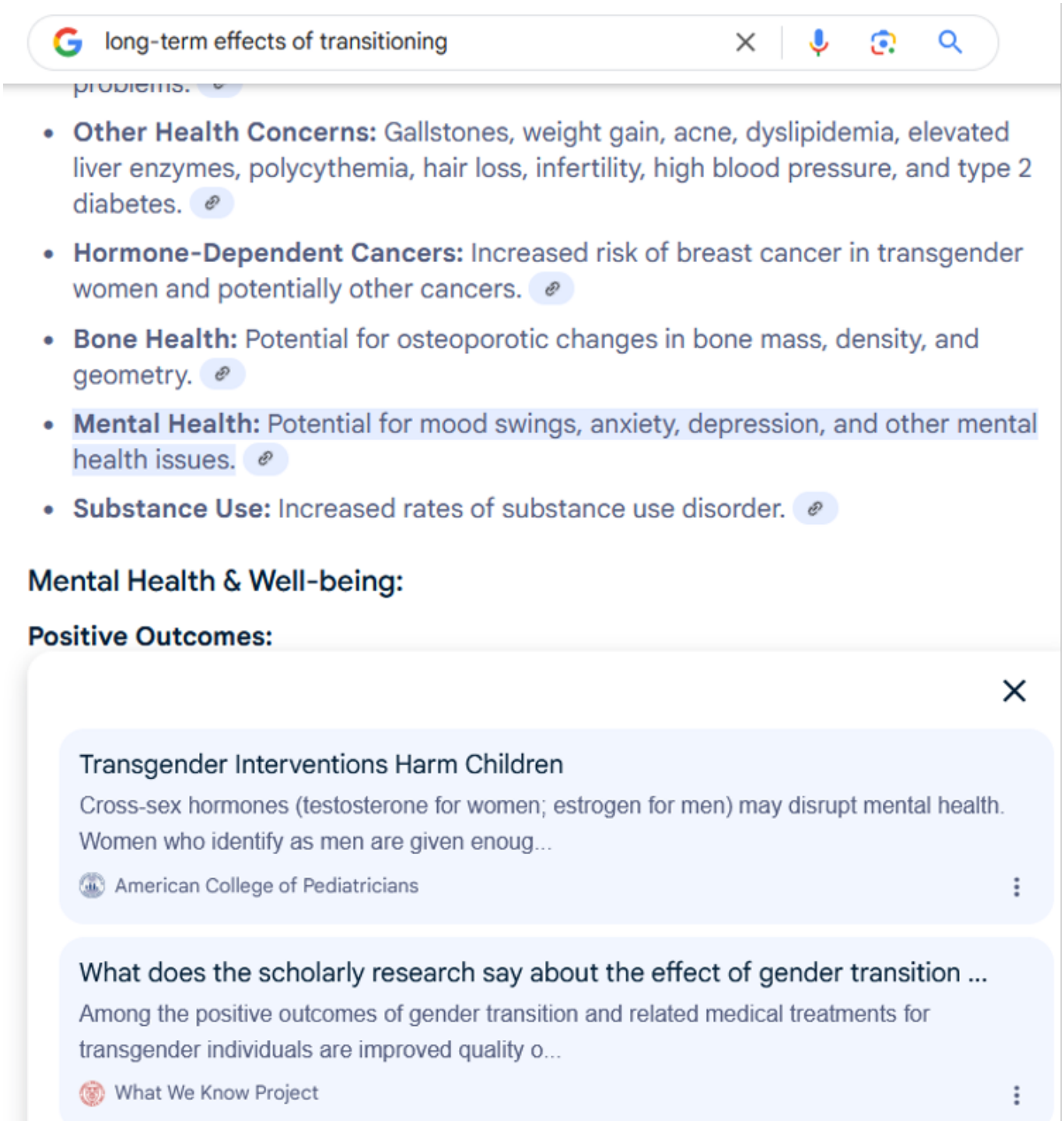


Figure 1. A screenshot of a Google Search query titled "long-term effects of transitioning." Below this query, an algorithmically generated text response is displayed. The sentence, "Mental Health: Potential for mood swings, anxiety, depression, and other mental health issues," is highlighted in blue to indicate that sourcing information is being provided. In the text box at the bottom of the screenshot providing sources, a blog post from the American College of Pediatricians, an anti-LGBTQ+ disinformation group, is featured, titled "Transgender Interventions Harm Children."

Here, I examine how disinformation campaigns can uniquely affect algorithmically generated summaries, shown through a case study of anti-transgender disinformation campaigns in the United States. Anti-transgender groups have increasingly attempted to deny transgender people's rights to healthcare, identity documents, and equal participation in sports, among other freedoms. In order for this reactionary movement to succeed, these groups need to popularize reasons to support anti-transgender policies like

these.[1] Their efforts to do so have included a broad, well-funded disinformation campaign to paint transgender people's presence in public life as problematic:[2] dangerous in public bathrooms, unfair in competitive sports, and undeserving of medical care. None of these claims are true, of course. Transgender people simply seek the right to a public life as any cisgender person might, but these campaigns have had worrying success. After a few years of seemingly expanded public rights for transgender people, we are now in the midst of a legal movement that is threatening to rescind them.

Setting aside the systematic impacts of the anti-transgender political movement, coming out and living as a transgender person is also a deeply personal and sometimes confusing process, both for the transgender person themselves and for their family, friends, and community. Previous research has documented how transgender people and their support systems use tools like search engines to seek out health-related information.[3] Tools like Google Trends, as well as Google's suggested searches for transgender-related queries, reveal a persistent interest in health information for transgender people. As the current anti-transgender disinformation campaign creeps into the new technology of algorithmically generated summaries, how can those seeking information about transgender people—often, transgender people themselves—be subtly entrapped by those aiming to mislead them?

The Algorithmic Laundering of Pseudoscience

A significant part of the US campaign against transgender public life is carried out through a *pseudoscientific disinformation campaign*. Many people in the US are familiar with such campaigns via encounters with antivaccine media, which typically (and falsely) claim that all or some vaccines are dangerous and/or ineffective. But pseudoscientific disinformation campaigns are as old as modern science itself, ranging from the advocacy of race science by slaveholders and their ideological descendants,[4] to the industry-funded campaigns to contest the link between smoking and cancer.[5]

Pseudoscientific disinformation campaigns have many tactics, but many of them rely on the concept of *scientific boundary work*. Sociologist Thomas Gieryn developed the term to describe how different, opposing groups of people work to draw the line between which information was "science," and which was "not-science," or unscientific.[6] Those wishing to contest the mainstream scientific perspective have twin challenges in front of them: How to make their own materials appear to be "science," and equally as important, how to make the mainstream, widely trusted science become "not-science." My own research on pseudoscientific disinformation campaigns shows how online influencers and organizations use both of these tactics to great effect, selectively and deceptively citing research to either tear down mainstream consensus, or elevate their own false consensus.[7]

In the context of transgender history, however, it is important to remember that not all efforts to contest the mainstream scientific consensus—to do boundary work—should be considered malicious or otherwise disinformation campaigns. Sometimes, these efforts represent corrections to a mainstream science whose perspectives were limited by institutional bias or a lack of relevant experience. For example, much of today's medical consensus for the provision of gender-affirming care is due to the hard-won boundary

work of transgender people and their allies in just the past few decades. The previous medical model restricted care to very few transgender people, often provided inadequate care for their needs, and was successfully characterized as "not-science" by transgender people who wanted broad access to competent healthcare.[8] As chronicled in a report by the Southern Poverty Law Center (SPLC),[9] our current moment of pseudoscientific backlash to transgender care is in fact supported by many discredited medical professionals who were once considered in the scientific "mainstream."

Of course, the classification of certain boundary work as a "disinformation" campaign is always contextual. As historian Naomi Oreskes describes in her book *Why Trust Science*, what has separated what could be called disinformation campaigns meant to corrupt science from genuine efforts to improve science is located not just in scientific failures, like poor methods or weak evidence.[10] The separator also lies in a given practitioner's personal values (and biases), their tolerance to opposing viewpoints of open critique, and ultimately their humility to the process of scientific contestation. When I refer to pseudoscientific disinformation in the context of anti-transgender movements, I do so to highlight those whose viewpoints have not only been rejected by their scientific and medical peers, but also whose personal values are opposed to transgender people's self-determination and free access to healthcare.

How does scientific boundary work show up in algorithmically generated search summaries? Software developers, though not always seen in the same light as scientists or organizers, engage in boundary work too—especially those making products for information-seeking. When users search for medical answers on their websites, and they claim to give them a credible answer, they must determine which sources their algorithm should trust, and which they should exclude. Google, for example, discussed its efforts to remove parody websites from its summaries, and stressed that it has "strong guardrails in place" for topics like health. In practice, these guardrails are boundary work, separating good medical knowledge from faulty medical knowledge that has no place appearing as legitimate health information in a search engine result.

Unfortunately, selecting high-quality information can be difficult in the best of times, and even more so in the face of a pseudoscientific disinformation campaign. Take, for example, information-seeking around the topic of "detransitioning." Anti-transgender movements have focused intense attention on the prospect of transgender people changing their public identification to cisgender ("detransitioning"), as a high rate of detransition could provide a basis for claims that transgender self-identification are either transient or the result of a curable mental disorder.[11] While current scientific evidence indicates that very few people detransition, pseudoscientific disinformation campaigns have suggested that the low mainstream estimates of detransition rates cannot be trusted due to methodological errors, and they have promoted limited studies suggesting the rate is much higher.

To understand how debate over this topic manifests in algorithmic summaries, we can look at the query "how many trans people detransition," which was one of the top ten queries related to the term "detransition" on Google Trends in the last five years (Figure 2). The summary begins with the main takeaway that "the rate of detransitioning among transgender individuals is estimated to be low, but there is no definitive consensus due to limitations in research." The sole source for this claim is

attributed to a blog post titled “Accurate Transition Regret and Detransition Rates Are Unknown,” which is authored by the Society for Evidence-Based Gender Medicine (SEGM), an organization the SPLC has called a “key hub of anti-LGBTQ+ pseudoscience.”[12] The rest of the summary proceeds similarly, with individual studies highlighted and then criticized on the basis of supposed methodological flaws.

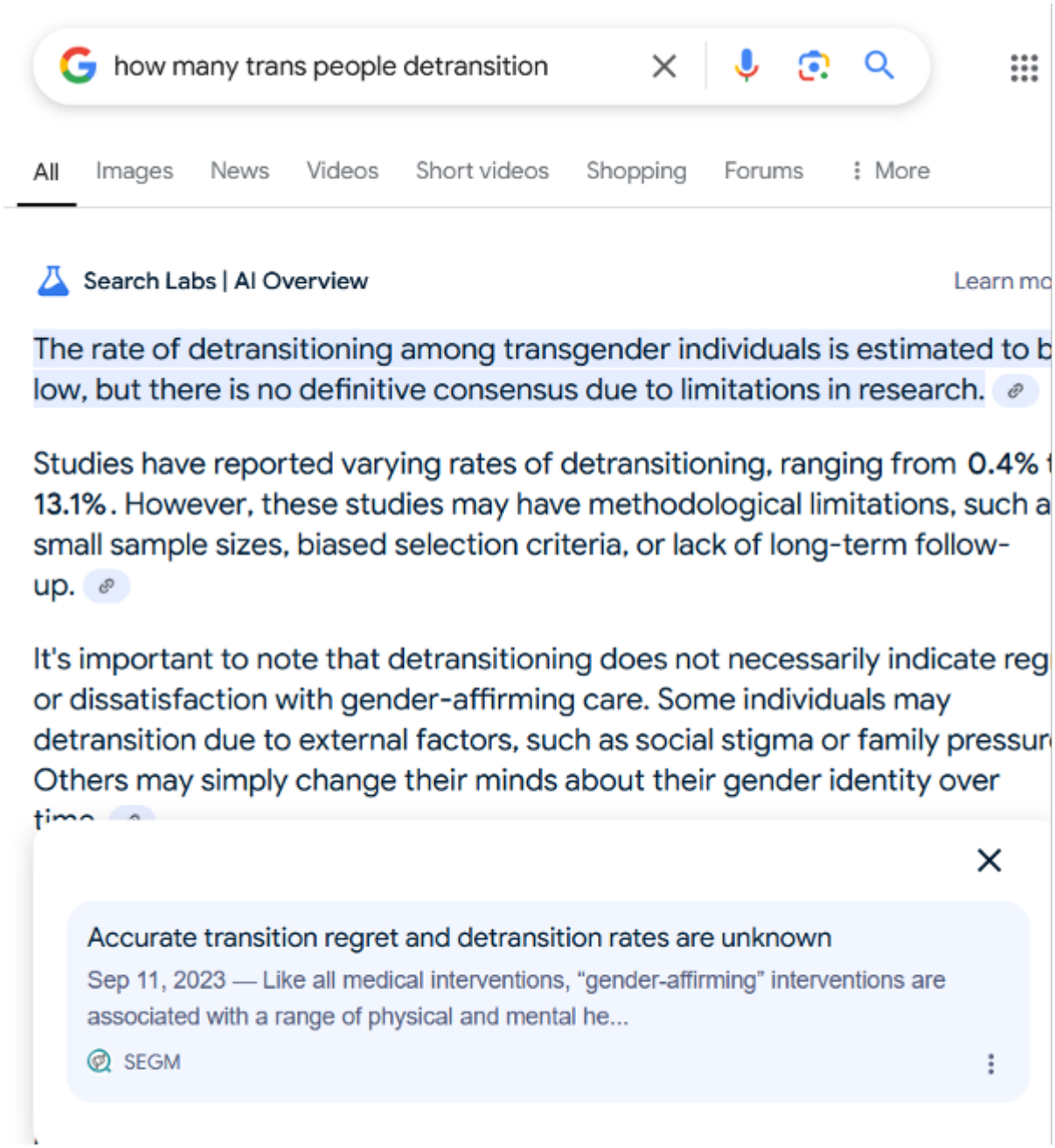


Figure 2. A screenshot of a Google Search query titled “how many trans people detransition.” Below this query, an algorithmically generated text response is displayed. The first sentence, “The rate of detransitioning among transgender individuals is estimated to be low, but there is no definitive consensus due to limitations in research,” is highlighted in blue to indicate that sourcing information is being provided. In the text box at the bottom of the screenshot, a blog post on the Society for Evidence-Based Gender Medicine (SEGM)’s website is highlighted, titled “Accurate Transition Regret and Detransition Rates Are Unknown.”

One might imagine that algorithms like these could improve their answers to medical questions if they restricted themselves to citing the academic literature, rather than potentially partisan blogs—that is,

sticking to the "science." However, pseudoscientific disinformation campaigns are often laundered through the scientific literature itself, creating additional challenges for algorithmic systems hoping to do scientific boundary work. Prior reporting has documented how SEGM pays for open-access fees for authors publishing research congenial to their anti-transgender stances, so that their work will be freely accessible to the public instead of hidden behind paywalls. This also means, however, that the text of these articles is available to algorithmically generated search summaries like Google Search, which often cites their claims via publicly accessible text.[13] Two out of the three peer-reviewed scientific articles cited in the above summary on detransition had their open-access fees paid for by SEGM, of these, one suggests that existing research on detransition rates is flawed, and the other suggests it is higher than is widely reported. One of the articles was authored by someone whose only listed affiliation was SEGM itself.

It should be noted that one of the most vexing things about studying disinformation in algorithmically generated summaries is how much they can change based on the slightest difference in the original search query. For example, Google's answer to the query "detransition rate" is currently quite different from its answer for the query "detransition rates," and provides entirely different percentages and cites different sources. This changeability can provide a certain plausible deniability for the developers of algorithms, who can always claim that flaws found in specific examples do not necessarily generalize to the wide diversity of queries used in the "real" world. To be assured in this preliminary analysis, I tested 20 additional queries relating to the rate of detransition. Google provided summaries in 18 out of 20 queries, and sources that are currently or were previously sponsored by SEGM were found in every summary.

The ultimate effect of Google's algorithmic citation decisions is to manufacture doubt on the present-day consensus on transgender care: that detransition is rare. While this algorithmically generated summary claims there "is no definitive consensus" on the frequency of detransition, this claim originates from content generated by or sponsored by anti-transgender partisan groups. This is ultimately a failure of the algorithms' developers to perform scientific boundary work, denying the mantle of scientific authority to anyone ("no definitive consensus") despite compelling reasons to give that authority to certain sources. As long as algorithmically generated summaries are asked to make claims on scientific consensus, and as long as they are unable to parse deceptive sources from mainstream consensus sources, they will be vulnerable to laundering claims that originate in pseudoscientific disinformation campaigns.

The Algorithmic Sanitizing of Bigotry

Communications researchers frequently study how members of the media *frame* information: how they choose words, imagery, or other media to emphasize certain aspects of a story, while minimizing others. Frames can sometimes be obvious, but often they are communicated through subtle word choices or omissions of context, and these choices can have societal consequences. Prior research, for example, has experimentally tested how the media's framing of an incident of police brutality toward a Black woman influenced readers' favorability toward the police officer involved.[14] When reported with a frame stressing law and order, such as describing what laws may have hypothetically been broken, support for

the officer increases, whereas when reported with a frame stressing police violence, such as the injuries sustained by the victim, support decreases. When different media report on the same event in different ways, people's perceptions of society's problems and who is to blame change accordingly.

In the case of generative search engines and large language model chatbots more generally, software developers have tried to make their algorithms appear "helpful, honest, and harmless," and have them frame issues in a nonpartisan manner. This built-in preconditioning affects the word choices that these algorithms use when writing responses to people's searches, subtly changing their answers' meanings. To understand how this tendency toward "harmless" and nonpartisan frames can propagate disinformation campaigns, we can take a look at another group involved in anti-transgender disinformation campaigns: anti-transgender Christian advocacy organizations.

Take, for example, the query "is my child trans or confused," a phrasing that is currently recommended via Google's autocomplete feature from the base query "is my child trans" (Figure 3). As of writing, the algorithmically generated summary cites eight sources, six of which are mostly supportive of children's right to identify themselves as transgender. The remaining two sources are blog posts from Christian organizations that resolutely oppose the notion that children be allowed to identify their own gender. The first Christian organization, Focus on the Family, has claimed that transgender identification is "a result of sin and deception entering the world," while the second, Christian organization, Unleash the Gospel, describes in their cited post that a particular child choosing to identify as transgender is "steeped in lies, self-hatred, and self-sabotage." Unsurprisingly, the blog posts referenced in this answer urge parents to resist and suppress their transgender children's self-identification, and both organizations recommend to parents what is commonly considered conversion therapy for transgender children.[15]

In earlier versions of search engines, which typically listed websites, users may be able to quickly identify context clues that could help them judge these anti-transgender websites for themselves. They may be familiar with Focus on the Family, which has advocated for anti-LGBTQ+ practices like homosexual conversion therapy for decades. They may be sensitive to the word choice of these articles, whose references to prayer and sin could signal a Christian ideology that is sometimes hostile to LGBTQ+ people. In the algorithmically generated summary, however, these signals are suppressed. The names of sources are more marginal in the presentation of search results, with the blog post from Focus on the Family only visible after scrolling downward on a sidebar. But more importantly, the algorithm paraphrases the language from these organizations to make them more palatable to a general audience: "helpful, honest, and harmless," as well as nonpartisan and nonreligious.

Take the two sections of this response attributed to the blog post from Unleash the Gospel. At different parts of the summary, the Unleash the Gospel blog post is cited for the claims "if you are worried about your child's feelings regarding their gender, reach out to a therapist who specializes in gender identity issues" and "don't rush to label your child as trans without seeking professional guidance." The author of the original Unleash the Gospel blog post recommends parents seek a therapist with a Christian perspective in order to reverse their child's transgender identification, i.e., transgender conversion therapy. But in the context of this algorithmically generated summary, the advice is ambiguous. Stripped of reference to Christian opposition to transgender identity, it may even seem helpful, honest, and

harmless: What child could not benefit from some therapy? But in the historical context of transgender care for children, it reinforces a model where transgender people lack the agency to identify themselves as such, and rather must be verified by third parties, like therapists. A significant part of the history of the struggle for transgender people to have the same rights as cisgender people is the right to seek healthcare without the permission of a therapist, and research reports that medical gatekeeping is a significant barrier to transgender people's access to healthcare.[16]

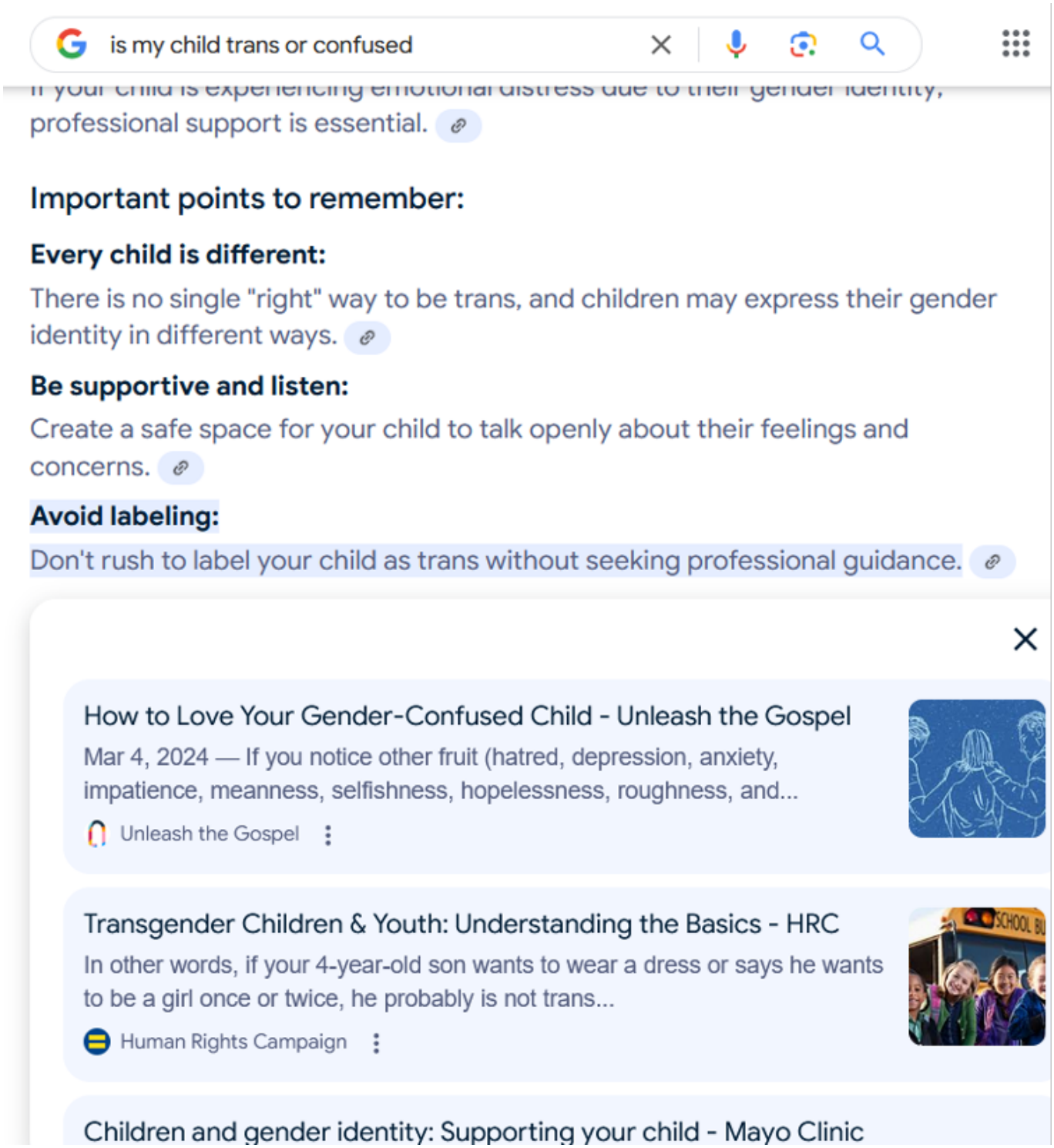


Figure 3. A screenshot of a Google Search query titled "is my child trans or confused." Below this query, an algorithmically generated text response is displayed. Midway through the response, a section with header "Avoid labeling" and the body text "Don't rush to label your child as trans without seeking professional guidance," is highlighted in blue to indicate that sourcing

information is being provided. In the text box at the bottom of the screenshot, multiple sources are provided, but first in the list is a blog post by Christian organization Unleash the Gospel titled "How to Love Your Gender-Confused Child."

Unlike in the previous case study on pseudoscientific disinformation campaigns, this particular vulnerability stems from *how* information from sources is conveyed, rather than *which* sources are selected, although both are problematic. Of course, those committed to transgender people's well-being would prefer that anti-transgender groups endorsing conversion therapy not be cited as sources in algorithmically generated summaries on parenting transgender children. But it is doubly harmful for a summary to not only cite these sources, but to make framing decisions that sanitize these sources' content for a general interest audience. By removing context clues for these sources, such as religious language or the actual reasons they advocate "therapy," the developers of these algorithms risk applying an indiscriminate normalization to deceptive practices that are ultimately harmful to transgender people.

One final factor to note is that the inclusion of these anti-transgender sources may be influenced by the phrasing of the query, as algorithmic chatbots tend to mimic the tone of their human conversation partners. For example, Microsoft, a developer of algorithmic search tools, has previously recommended that users be "polite" with chatbots, so that they may mirror that politeness in return. In this case, someone who asks the question "is my child trans or confused" may, through their choice of the word "confused," already betray some skepticism toward transgender identity, as gender "confusion" is not a term typically used in trans-supportive literature and communities. Google's algorithmically generated summary could be mirroring the skepticism from its user's query: both anti-transgender Christian websites in this summary reference "gender confusion" in their advice to suppress children's transgender identity.

While further research is needed to understand how search engines' users' pre-existing beliefs affect the algorithmic summaries they receive, previous research has shown this effect at work in standard, website-ranking search engines. For example, research has shown when opponents to abortion use Google search to learn more, they tend to choose queries that surface (sometimes misleading) websites opposed to abortion.[17] Information retrieval systems have always struggled with confirmation bias in this way, where "doing one's own research" involves merely seeking confirmatory (and sometimes low-quality) information to support an existing perspective.[18] Algorithmically generate search summaries may be providing new, complex ways for people to fall into this trap of self-confirming information.

Future Research on Disinformation in Algorithmic Search

These two case studies on anti-transgender disinformation illustrate a central point: Algorithmically generated search summaries will change—indeed, is already changing—the way misleading information is propagated on the internet. Software companies' decision to allow algorithms to make direct truth claims to search engine users makes them responsible for sorting fact from fiction in the midst of active disinformation campaigns—a battle they will often lose, even if they care to fight it. And their commitment to making algorithms appear authoritative and unbiased, while surely useful in some cases,

leads them to frame harmful and deceptive claims as normal, and even helpful.

It should be noted in both of these cases that they are not necessarily inherent to the technology of algorithmic text generation, but they are the consequence of these companies' empowering these often-incorrect algorithms to provide us with safe information about our social worlds. While it is doubtless that technical improvements to these algorithms' summarization ability can be made, the decision to allow the propagation of disinformation campaigns in algorithmically generated search summaries is ultimately one consciously made by the developers of this technology. Addressing disinformation in these summaries thus involves not only methodological changes to search features but advocacy efforts aimed at these developers, who can choose when and how to deploy a technology that has the capability to empower anti-transgender and other reactionary social movements.

Footnotes

1 Thomas J. Billard, "The Politics of Transgender Health Misinformation," *Political Communication* 41, no. 2 (2024): 344–352, https://doi.org/10.1080/10584609.2024.2303148.

2 R.G. Cravens, ed., *Combating Anti-LGBTQ+ Pseudeoscience* (Southern Poverty Law Center, 2023).

3 Laima Augustaitis et al., "Online Transgender Health Information Seeking: Facilitators, Barriers, and Future Directions," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York: Association for Computing Machinery, 2021), 1–14, https://doi.org/10.1145/3411764.3445091; Yolanda N. Evans et al., "Understanding Online Resource Use by Transgender Youth and Caregivers: A Qualitative Study," *Transgender Health* 2, no. 1 (2017): 129–139, https://doi.org/10.1089/trgh.2017.0011; Blake Hawkins and Jen Jack Gieseking, "Seeking Ways to Our Transgender Bodies, by Ourselves: Rationalizing Transgender-specific Health Information Behaviors," *Proceedings of the Association for Information Science and Technology* 54, no. 1 (2017): 702–704, https://doi.org/10.1002/pra2.2017.14505401122.

4 Nell Irving Painter, *The History of White People* (W. W. Norton & Company, 2010).

5 Naomi Oreskes and Erik M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (Bloomsbury Press, 2011).

6 Thomas F. Gieryn, "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists," *American Sociological Review* 48, no. 6 (1983): 781–795, https://doi.org/10.2307/2095325.

7 Anna Beers et al., "Selective and Deceptive Citation in the Construction of Dueling Consensuses," *Science Advances* 9, no. 38 (2023) eadh1933, https://doi.org/10.1126/sciadv.adh1933.

8 Carolyn Wolf-Gould et al., eds., *A History of Transgender Medicine in the United States: From Margins to Mainstream* (State University of New York Press, 2025).

9 "Foundations of the Contemporary Anti-LGBTQ+ Pseudoscience Network," in *Combating Anti-LGBTQ+ Pseudeoscience*, ed. R.G. Cravens (Southern Poverty Law Center, 2023).

10 Naomi Oreskes, "Science Awry" in *Why Trust Science?* (Princeton University Press, 2021). 127–139.

11 Joanna Wuest and Briana S. Last, "Agents of Scientific Uncertainty: Conflicts over Evidence and Expertise in Gender-affirming Care Bans for Minors," *Social Science and Medicine* 344 (March 2023) 116533, https://doi.org/10.1016/j.socscimed.2023.116533.

12 "Group Dynamics and Division of Labor within the Anti-LGBTQ+ Pseudoscience Network," in *Combating Anti-LGBTQ+ Pseudeoscience*, ed. R.G. Cravens (Southern Poverty Law Center, 2023).

13 Laura Acion et al., "Generative AI Poses Ethical Challenges for Open Science," *Nature Human Behavior* 7 (2023): 1800–1801, https://doi.org/10.1038/s41562-023-01740-4.

14 Kim Fridkin et al., "Race and Police Brutality: The Importance of Media Framing," *International Journal of Communication* 11 (2017): 21–21, https://ijoc.org/index.php/ijoc/article/view/6950.

15 Florence Ashley, "Interrogating Gender-Exploratory Therapy," *Perspectives on Psychological Science* 18, no. 2 (2023): 472–481, https://doi.org/10.1177/17456916221102325.

16 Meaghan B. Ross et al., "Voices from a Multidisciplinary Healthcare Center: Understanding Barriers in Gender-Affirming Care—A Qualitative Exploration," *International Journal of Environmental Research and Public Health* 20, no. 14 (2023): 6367, https://doi.org/10.3390/ijerph20146367.

17 Francesca Bolla Tripodi and Aashka Dave, "Abortion Near Me? The Implications of Semantic Media on Accessing Health Information," *Social Media + Society* 9, no. 3 (2023), https://doi.org/10.1177/20563051231195548.

18 Francesca Bolla Tripodi, Lauren C. Garcia, and Alice E. Marwick, "'Do Your Own Research': Affordance Activation and Disinformation Spread," *Information, Communication and Society* 27, no. 6 (2024): 1212–1228, https://doi.org/10.1080/1369118X.2023.2245869.