Weird Science: Scientific Research is Threatened by Much More Than Just AI Slop

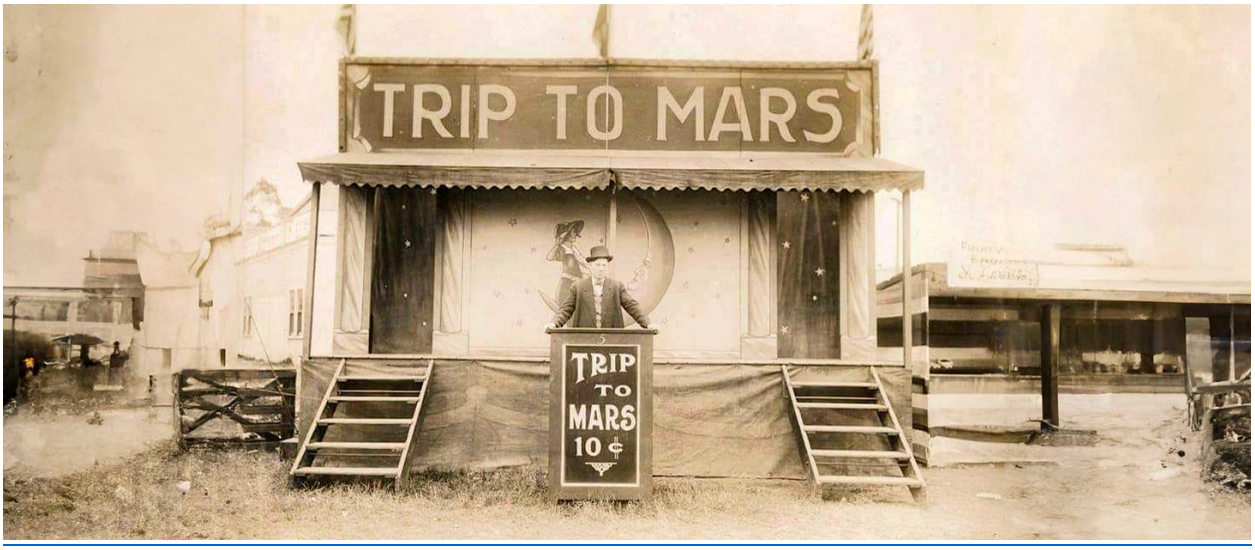
Roberta Munoz

New York University

ABSTRACT: This piece talks about some of the hidden harms of integrating generative AI tools into academic research. The dangers go beyond fake citations and confabulated facts. GenAI infuses the entire research process with distortions and biases. I specifically talk about the MAHA report, and the use of AI powered “research assistants” like Google NotebookLM. NotebookLM treats all sources equally positively, and sycophantically; even studies that have been debunked or are fraudulent. Google's statement that it's entirely "grounded" in your own sources is not accurate. This is because the LLMs that underly all functions of this AI powered tool were created and trained to power commercial products like chatbots. All "derivatives" (summaries, overviews) that NotebookLM creates are filtered through training layers and content moderation layers to promote polite engagement. GenAI creates probabilistic and algorithmic connections that undermine traditional and stable ways of creating scientific consensus. The "Key Topics" attached to sources by NotebookLM refer to no known, or stable, knowledge organization framework. These concepts and keywords are generated from the sea of chaos that is an LLM - blogs, social media, junk science. I tested multiple junk science studies, and I use a notorious bad researcher name Brian Wansink as a prime example. Wansink has the dubious distinction of having had six scientific articles retracted in a single day! And yet, NotebookLM summarizes his work most favorably. LLM model and their filters designed for social chatbots are now influencing academic research where studying sensitive and controversial topics is essential.

Keywords: generative AI, academic research, GenAI, AI enhanced discovery, AI enhanced search, AI tools

Journal of Radical Librarianship, Vol. 12 (2026) pp.25-28. Published 15 April 2026.



Of course the MAHA (Make America Healthy Again) report released earlier this year cited non-existent sources. As an academic librarian who has seen a tsunami of AI-fabricated citations from students (and faculty) during the last two years I was surprised/not surprised.

Did they use AI tools to discover and retrieve the citations? Probably. Fake research studies are a serious problem in the age of Generative AI. But the dangers go way beyond citation slop. GenAI infuses the entire research process with subtle distortions and biases. Worse, it decontextualizes scientific findings and undermines traditional and stable ways of creating scientific consensus. There's evidence of this throughout the text of the MAHA document.

The made-up citations in the MAHA paper were first reported by NOTUS. In addition to the fabricated citations, the NOTUS also found overgeneralization of legitimate studies and misleading emphases of findings. The study also found a tendency for AI to extrapolate scientific results beyond the claims found in the material that the models summarize. That is, it introduced concepts, keywords, and ideas in the derivatives that it created that were not present in the source material. How does this happen and why?

It's Bad Science all the Way Down

I'll use Google NotebookLM as an example here. NotebookLM is advertised as a research

organization tool. It lets you upload sources and notes and then 'synthesizes' the information into derivative artifacts like summaries, "mind-maps", audio overviews (in the form of an imitation podcast), and now video overviews.

NotebookLM was created to be a research assistant and a research organization tool modeled on traditional tools like ZOTERO, RefWorks, and Endnote. The selling point, according to Google is that, because NotebookLM uses "source-grounding" (Google's term), i.e. generating materials only from your own uploaded sources, the outputs are more accurate, objective, and less likely to contain hallucinations. (see the product description here). This is not true.

Hallucinations are here to stay and NotebookLM hallucinates, sometimes. All LLMs have biases; there is no objective 'view from nowhere', and most LLMs have a pronounced bias is towards positivity and sycophancy. A version of ChatGPT released recently by OpenAI was so unctuously agreeable that it caused a massive backlash. It was later toned down but the effect didn't disappear entirely. Flattery is part of the DNA of GenAI as you can see when browsing system prompts for various models (information here).

When using AI-enhanced 'research assistants' like NotebookLM however, the problem is deeper than just a cringe-worthy conversation with a chatbot. A science educator put it this way when she said "It's not just that AI is too nice to the user, it's *too nice to nonsense*. And it's true. NotebookLM is *way* too OK with bad, questionable, and garbage research.

NotebookLM doesn't just treat good and bad sources *equally* - it treats them *equally positively*. Those podcasters in the audio overview have never met a research paper they didn't like.

The reason: every AI-generated derivative is deeply enmeshed in how the underlying model operates. NotebookLM, and other tools like it (OpenAI Deep Research for example) all generate outputs within the hidden "soft cage" of LLM protocols. The parameters, added during model training, ensure that every word is unobjectionable, that contradictions are minimized or erased, and unsettling truths are either absent or wrapped up tightly in reassurance.
This is much more destructive to the research process than even AI citation slop. Bad research is presented either neutrally or positively, even when studies have been retracted and identified as fraudulent.

I did some testing by uploading some the worst, most debunked, and frequently retracted research papers I could find, and I have to say that NotebookLM was *extremely* OK with all of it.

The summaries were upbeat. And the audio overviews? I just can't with these podcasters. When discussing the most toxic and destructive scientific ideas of the last few decades, they were absolutely jazzed. What?

Let's look at an article published by a notorious scam scientist.

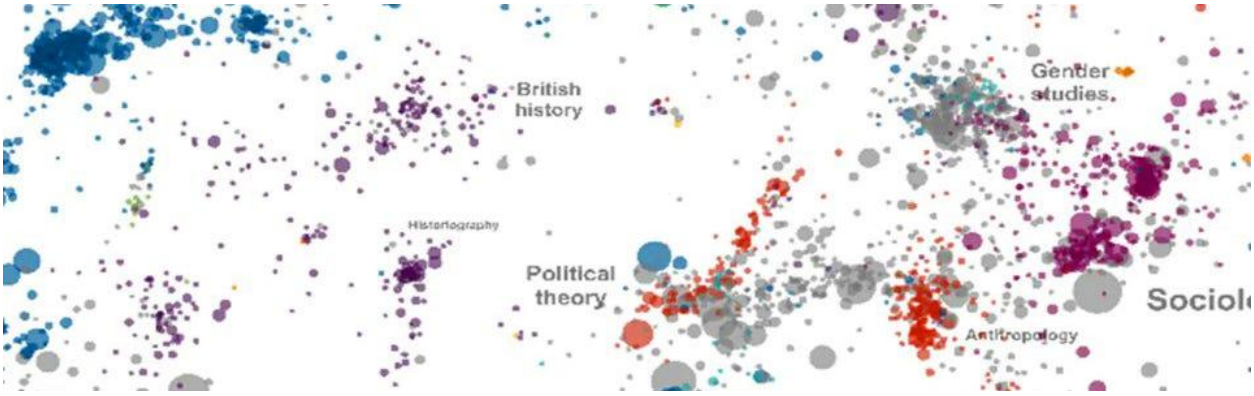
Brian Wansink was a master of junk science who has the unique distinction of having had 6 of his "studies" retracted *in one day*. He's the author of a fake study called "The Consequences of Belonging to the Clean Plate Club." The ideas in this article were widely discussed and accepted from 2008 until the paper was retracted in 2017, almost 9 years later.

The first AI audio overview that I generated based on this "study" was massively positive. The 'podcasters' used frequent affirming language and expressions, like "makes sense", and "yes, I understand". Ideas present in the source material were enthusiastically confirmed without a hint of skepticism. New ideas like "internal regulation" and "helicopter parenting", not present in the source material, were introduced to support the now disgraced study. Uploading the same source multiple times resulted in different content within the derivatives, and all were presented with either uncritical neutrality or extreme enthusiasm.

Why were new concepts and keywords not present in the original source generated, and generated anew and differently with every upload? Because underneath the "polished layer" of every LLM model that's refined with RLHF (Reinforcement Learning from Human Feedback), RAG (Retrieval-Augmented Generation), content moderation guidelines, and AI charm school, lies something else.

Despite Google's claims of 'grounding' their outputs only in your sources, all of NotebookLM's synthesized derivatives tap into and express the underlying, chaotic mass of unstructured data that serves as the primary fuel for every LLM engine. When generating outputs, AI tools by their very nature crack the fragile surface of their training layer and bring up data from the blogs, social media posts, rage bait, rants, wish fulfillment, and fantasy that make up the model's substratum.

Sycophancy is built into the top of most LLMs, but this positivity bias is also a result of this subterranean data. The ideas in "The Clean Plate Club" felt true at the time, and so were widely disseminated throughout the media and the info-sphere for nine years. AI has ingested this data and recirculated and reconstituted it, even though the study was ultimately exposed as trash. This article was originally published in a peer-reviewed journal. When retrieved from a database, it comes with connected concepts in the form of the subject headings and taxonomies that create a stable, human-created framework used by scholars to situate novel research within the field. In this case, MeSH subject headings and citation histories are present. This information is stripped and absent when this, or any other piece of research, is uploaded into NotebookLM. The only context that remains is AI-generated - fashioned from the digital detritus that fuels it.



We teach another class at my university called “Mapping the Field”. To do meaningful scientific research you have to get to know the field. So, what makes a field a field? Our definition:

• Coherence: shared research agenda(s), methodology

• Authorities

• Terminology/discourse

• Structure

These create a map with agreed-upon boundaries (structure), a common language (terminology/discourse), and stable guideposts.

GenAI tools can make maps and create language, but the words that surface as "Key Topics" in Notebook are derived from the uploaded sources in isolation and are disconnected from any system of knowledge organization. Terms and concepts that emerge from the hidden layers of the LLM are equally disembodied. Notebook’s research path has no guideposts and only an illusion of progress. It acts like an infinitely regenerating road to nowhere in some strange video game where, with every step you take, a new world recreates itself around you, directionless and unconnected to anything but itself.

Traditional terminology and categories, scientific paradigms, can be limiting, but they are necessary for the creation of a coherent and shared scientific conversation. GenAI tools, NotebookLM included, break this method of creating scientific consensus. You don’t know where you are on the map. There is no clear way forward. But it’s all good, right?