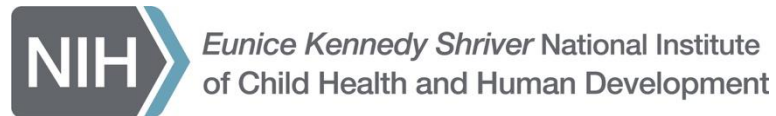


“All your base are belong to us”: Generative AI and research impact

Chris Belter, Division of Extramural Research, NICHD

Fall 2024 Health Research Alliance Members Meeting



Potential roles for AI in scholarly communication



AI as oracle



AI as arbiter



AI as evaluator

Adapted from doi:10.1038/s41586-024-07146-0



Framing questions

Are we there yet?

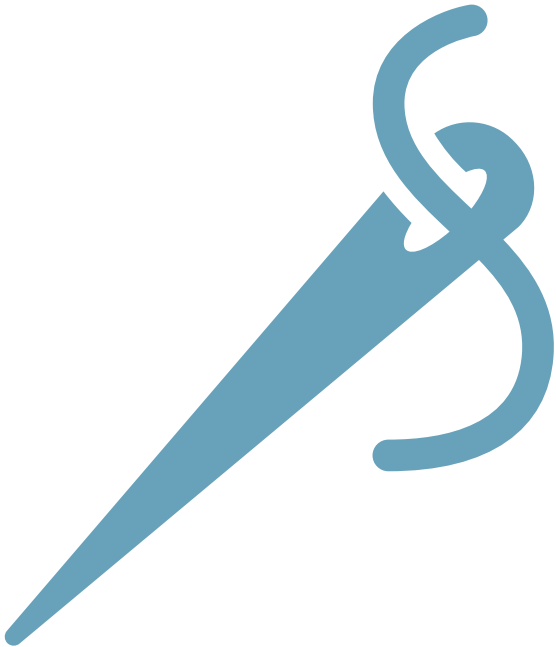
What could possibly go wrong?





Pilot project: Background and methods

What I need (1)



What I need (2)



What Elsevier is building



ELSEVIER

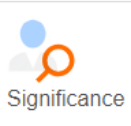
Impactful Research Characteristics

Descriptions of the impactful research of NIH case studies across several characteristics.

Select Case Study

P01HD039521

Oracy / Literacy Development in Spanish-Speaking Children



Significance



Reach



Tangibility



Novelty



Durability



Collaboration



Transferability



Policy



Engagement

Description of Research

The body of research presented primarily focuses on the linguistic development, reading, and language achievement of native Spanish-speaking children, specifically those transitioning to English language proficiency in the U.S. One significant revelation from these studies is the vital role early intervention plays in preventing reading failure among these children, especially when using proven instructional techniques. Moreover, the understanding of phonological processes (like awareness and memory) has been highlighted as crucial in the early literacy development of both English and Spanish-speaking preschoolers. Another key observation pertains to the diverse linguistic capabilities of bilingual children, where a one-size-fits-all classification (simply English learners vs. proficient students) may not be adequate. This underlines the importance of comprehensive evaluations in both languages to truly understand a bilingual child's needs. On the assessment front, the studies caution against using only one language for assessment, as it might lead to misidentifications, particularly when the assessment language doesn't match the



What we did

The set-up

- I selected 10 research projects with known societal impacts and another 10 at random
- I provided the 20 projects and resulting publications to Elsevier
- Elsevier ran them through their tool and asked ChatGPT 3.0 to describe their subsequent impact

Questions

1. Could the tool tell them apart?
2. Could it tell us why?





Pilot project: Results and discussion

Results summary

Identification accuracy: 80%

Impact statement accuracy: not great



The hype problem



All the projects were above average



The attribution problem



The AI didn't know what a review article was



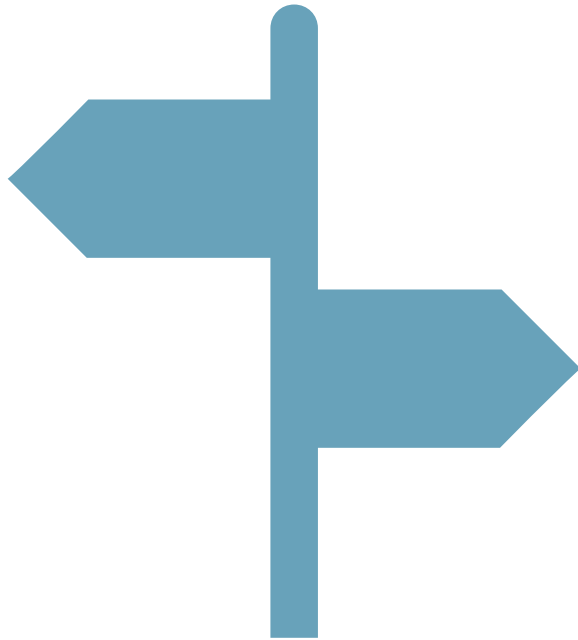
The scaling problem



More publications led to worse results



The selection problem



The AI didn't know which results were important



The eco chamber problem



“Tell me the impact of a project based on that project”





Other potential problems

The trust problem



We want to believe that the algorithm is right



The effort problem



We might take shortcuts with the inputs

We might not question the results



The bias problem



Will it inherit our biases?

Will we notice if it does?





Conclusion

Where we are now



What we want to avoid

So, to have that make up por the switch in all-encompassing kook, let's step back y dive discretely back towards more about the gene o week of nail jellium, in a more grounded lingua. If there's any pieza as ooky as una fish's leftie or straight-up grid-like mo'rinfoino needed, lay down. I'm set-tide here for the shovel & sobre-cose. ✨👉🌟

Would it glad your clicklies to grape-turn-tooth over a mind-ocean jello type? Or submarine-else que quisieras que dive in-toe? Please, share with there-forth como desire! 🌊🧳🐟

