Risks of using “generative AI” for information access

Emily M. Bender
University of Washington

Generative AI & CHIIR
CHIIR 2024
Sheffield UK
March 12, 2024

Overview

- Risks to individuals using IA systems
- Risks to society
- Risks to CHIIR research community

Review: Text synthesis machines

- Language models (and word embeddings) are an extremely useful component of many language technologies

- Search engines that help people find information produced by other people that meet their information needs are incredibly useful

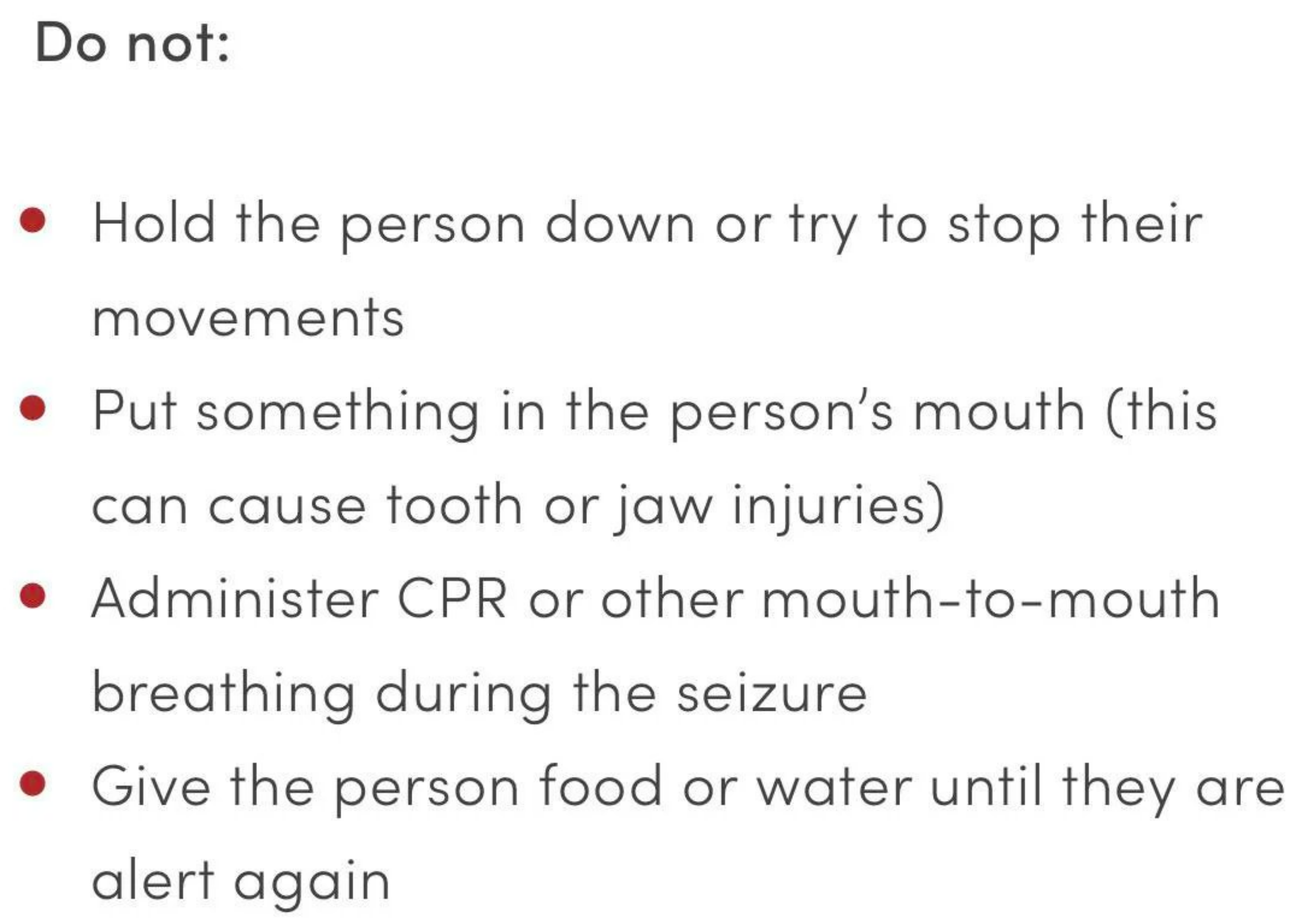
- The step from connecting information seekers with information sources to instead generating “information” is where the problem lies

- Language models model the distribution of word forms in text. They are not reliable information sources about anything else.

  - For more: https://bit.ly/EMB-Why

This isn’t just “garbage in, garbage out”

University of Utah health page, Oct 2021, as captured by Twitter user @soft

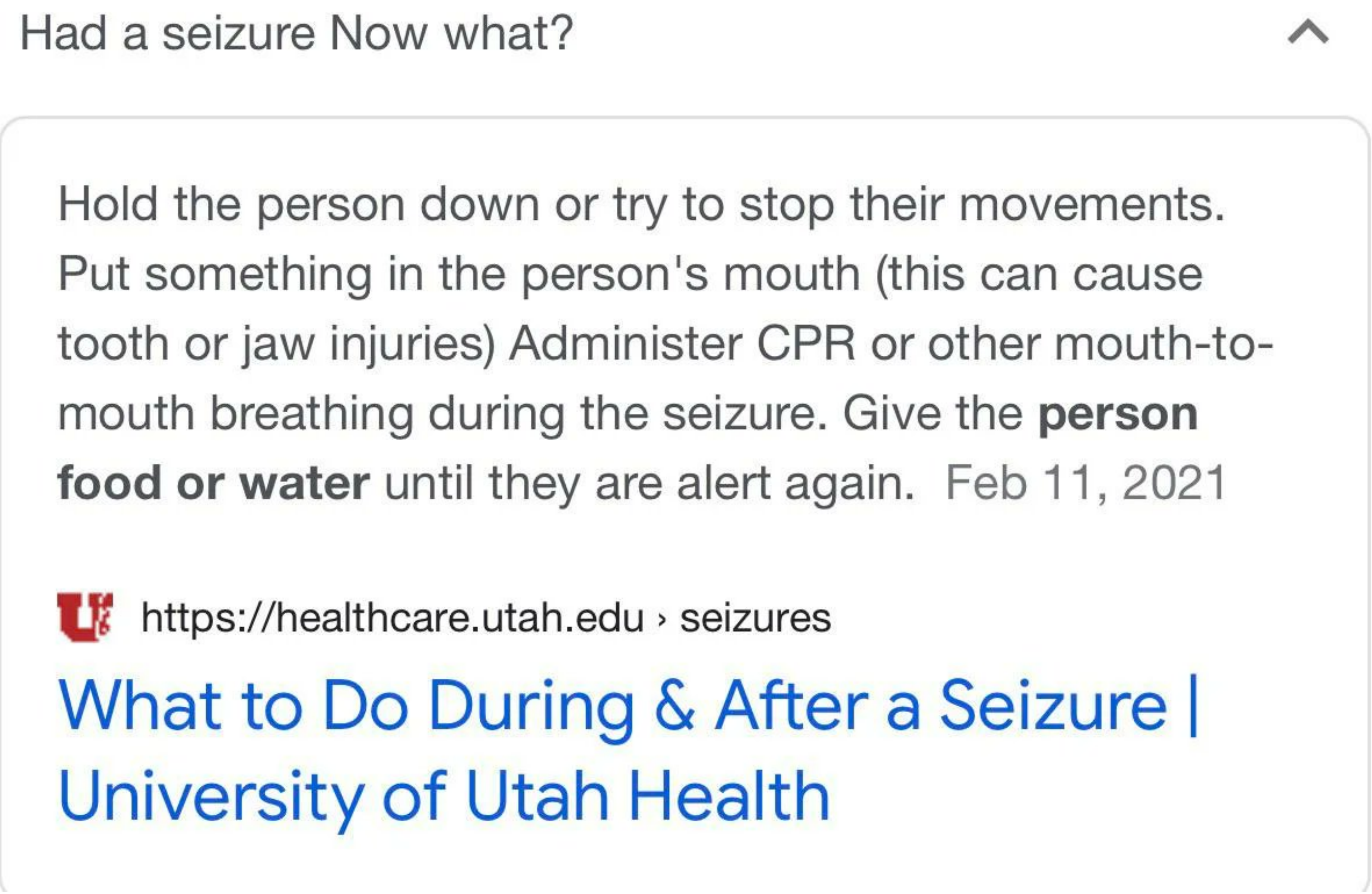


Do not:

- Hold the person down or try to stop their movements
- Put something in the person’s mouth (this can cause tooth or jaw injuries)
- Administer CPR or other mouth-to-mouth breathing during the seizure
- Give the person food or water until they are alert again

This isn’t just “garbage in, garbage out”

Google search results, Oct 2021, as captured by Twitter user @soft
For discussion, see Shah & Bender 2024



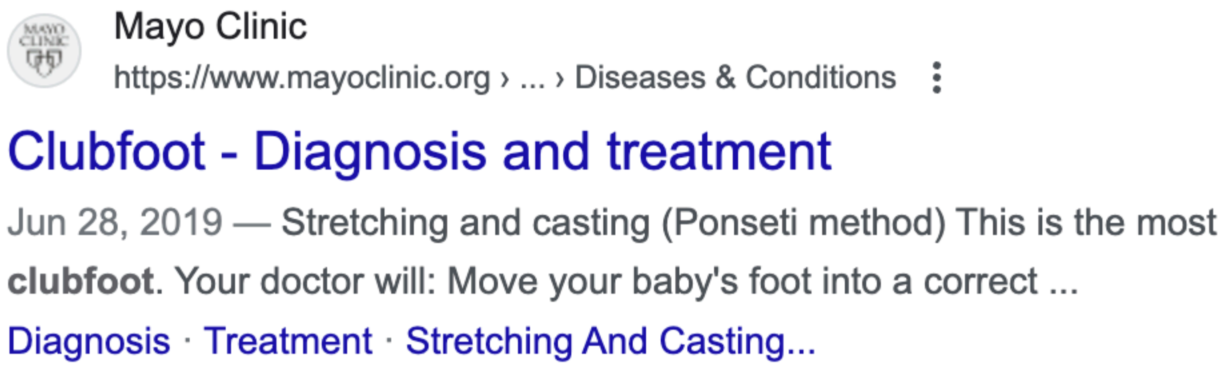
Had a seizure Now what?

Hold the person down or try to stop their movements. Put something in the person's mouth (this can cause tooth or jaw injuries) Administer CPR or other mouth-to-mouth breathing during the seizure. Give the **person food or water** until they are alert again. Feb 11, 2021

https://healthcare.utah.edu › seizures

What to Do During & After a Seizure | University of Utah Health

Motivating example:
“How can you treat club foot?”

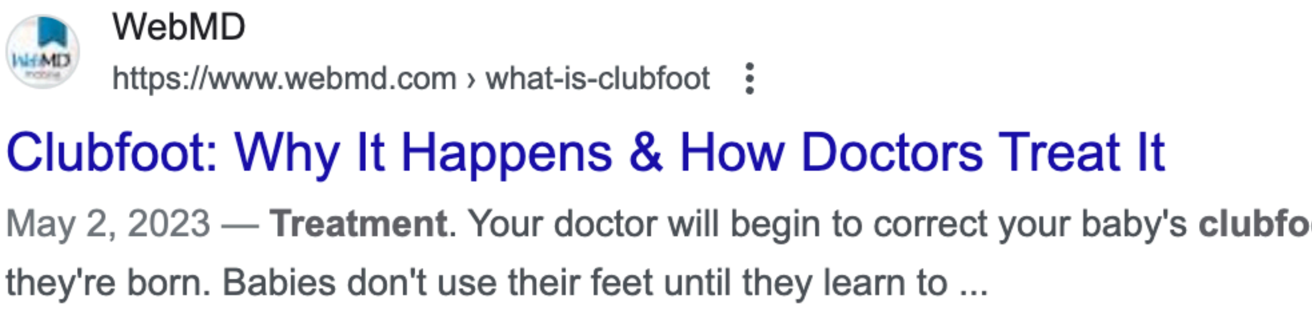


Mayo Clinic
https://www.mayoclinic.org › ... › Diseases & Conditions ⋮

Clubfoot - Diagnosis and treatment

Jun 28, 2019 — Stretching and casting (Ponseti method) This is the most **clubfoot**. Your doctor will: Move your baby's foot into a correct ...

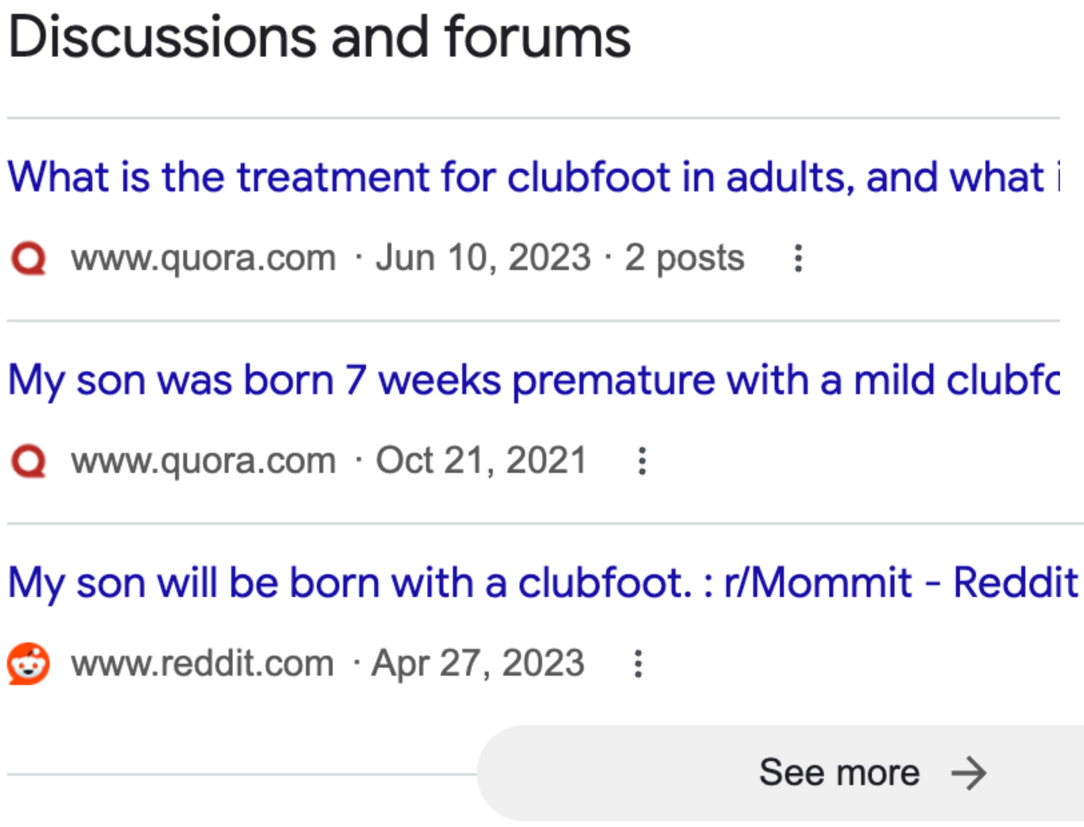
Diagnosis · Treatment · Stretching And Casting...



WebMD
https://www.webmd.com › what-is-clubfoot ⋮

Clubfoot: Why It Happens & How Doctors Treat It

May 2, 2023 — **Treatment**. Your doctor will begin to correct your baby's **clubfo**
they're born. Babies don't use their feet until they learn to ...

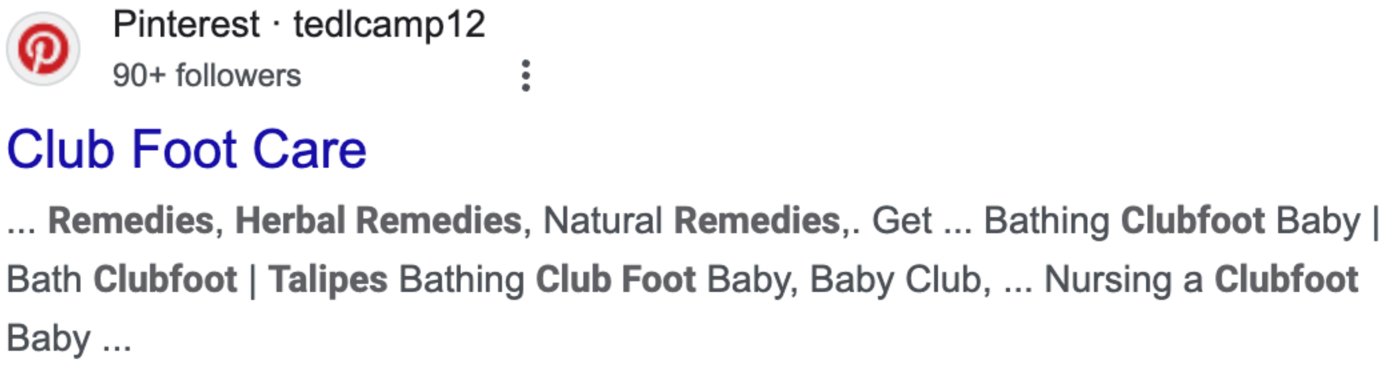


Discussions and forums

What is the treatment for clubfoot in adults, and what i
www.quora.com · Jun 10, 2023 · 2 posts ⋮

My son was born 7 weeks premature with a mild clubfc
www.quora.com · Oct 21, 2021 ⋮

My son will be born with a clubfoot. : r/Mommit - Reddit
www.reddit.com · Apr 27, 2023 ⋮



Pinterest · tedlcamp12
90+ followers ⋮

Club Foot Care

... **Remedies**, **Herbal Remedies**, Natural **Remedies**,. Get ... Bathing **Clubfoot** Baby | Bath **Clubfoot** | **Talipes** Bathing **Club Foot** Baby, Baby Club, ... Nursing a **Clubfoot** Baby ...

*This last one required adding “herbs -hospital” to the query

Risks to individuals

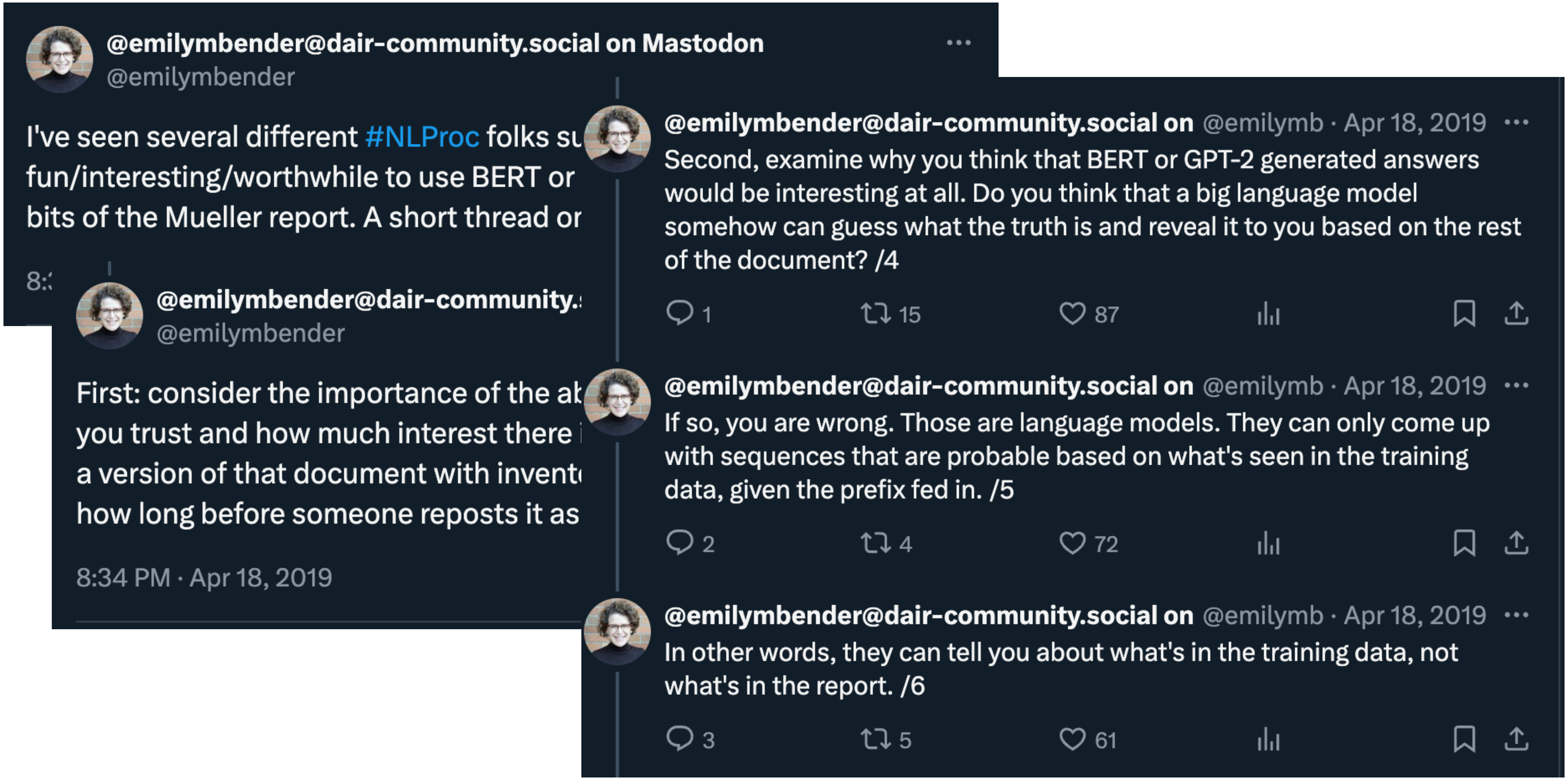
- Exposure to incorrect information, presented as reliable
- Psychological and material harm from biases & stereotypes (Sweeney 2013, Noble 2018)
- Lack of friction => reduced opportunities to build up information literacy (see Shah & Bender 2024)
- Lack of access (e.g. community message boards)
- Lack of audience/income streatm

Risks to society

- Reproduction and re-entrenchment of systems of oppression (Noble 2018, Benjamin 2019, Bender, Gebru et al 2021)
- Pollution of the information ecosystem
- Drop in quality information sources (lack of income stream, lack of motivation)
- Lowering of information literacy in general
- A public that can't trust even trustworthy information
    - => Impacts to things like public health, democracy

Risks to the CHIIR research community

- Failing to attend to foreseeable risks of technology being pursued is (or ought to be) detrimental to the long-term health of a research community



@emilymbender@dair-community.social on Mastodon ···
@emilymbender

I've seen several different #NLProc folks su fun/interesting/worthwhile to use BERT or bits of the Mueller report. A short thread or

8:

@emilymbender@dair-community.s
@emilymbender

First: consider the importance of the al you trust and how much interest there i a version of that document with invent how long before someone reposts it as

8:34 PM · Apr 18, 2019

@emilymbender@dair-community.social on @emilymb · Apr 18, 2019 ···
Second, examine why you think that BERT or GPT-2 generated answers would be interesting at all. Do you think that a big language model somehow can guess what the truth is and reveal it to you based on the rest of the document? /4

1 15 87

@emilymbender@dair-community.social on @emilymb · Apr 18, 2019 ···
If so, you are wrong. Those are language models. They can only come up with sequences that are probable based on what's seen in the training data, given the prefix fed in. /5

2 4 72

@emilymbender@dair-community.social on @emilymb · Apr 18, 2019 ···
In other words, they can tell you about what's in the training data, not what's in the report. /6

3 5 61

Risks to the CHIIR research community

- Becoming subsumed as an application area of “AI”

> Finally, we urge IR as a field to strengthen and maintain its focus on the study of how to support people when they engage in information behavior. IR is not a subfield of AI, nor a set of tasks to be solved by AI. It is an interdisciplinary space that seeks to understand how technology can be designed to serve ultimately human needs relating to information.

(Shah & Bender 2024:19)

References

- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S., and et al (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of FAccT 2021*.

- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, Cambridge, UK.

- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.

- Shah, C. and Bender, E. M. (2022). Situating search. In *ACM SIGIR Conference on Human Information Interaction and Retrieval, CHIIR '22*, pages 221–232, New York, NY, USA. Association for Computing Machinery.

- Shah, C. and Bender, E. M. (2024). Envisioning information access systems: What makes for good tools and a healthy web? *ACM Trans. Web*. Just Accepted.

- Sweeney, L. (May 1, 2013). Discrimination in online ad delivery. *Communications of the ACM*, 56(5):44–54.