What ethical challenges are looming with AI, and how will they be overcome?

IS3R 2023

Berlin/Germany

August 24–26, 2023

Adrian Brady

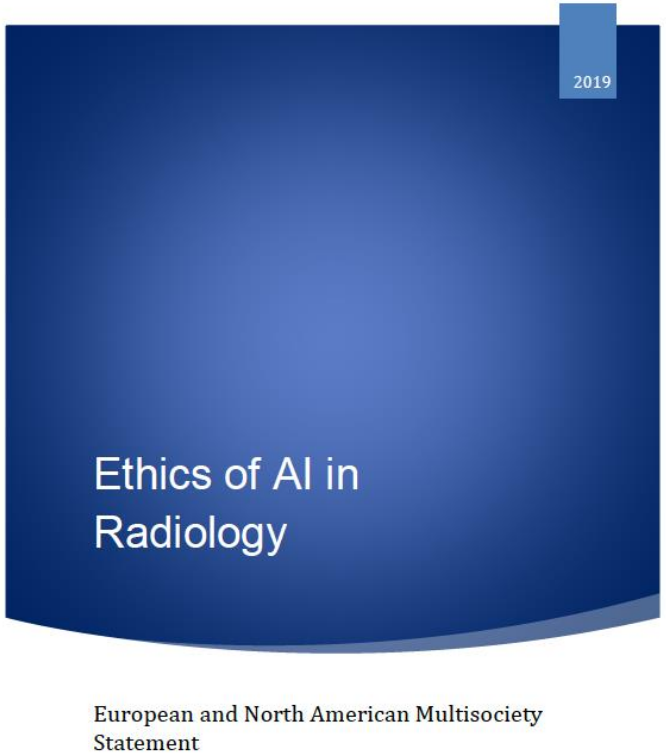
ESR

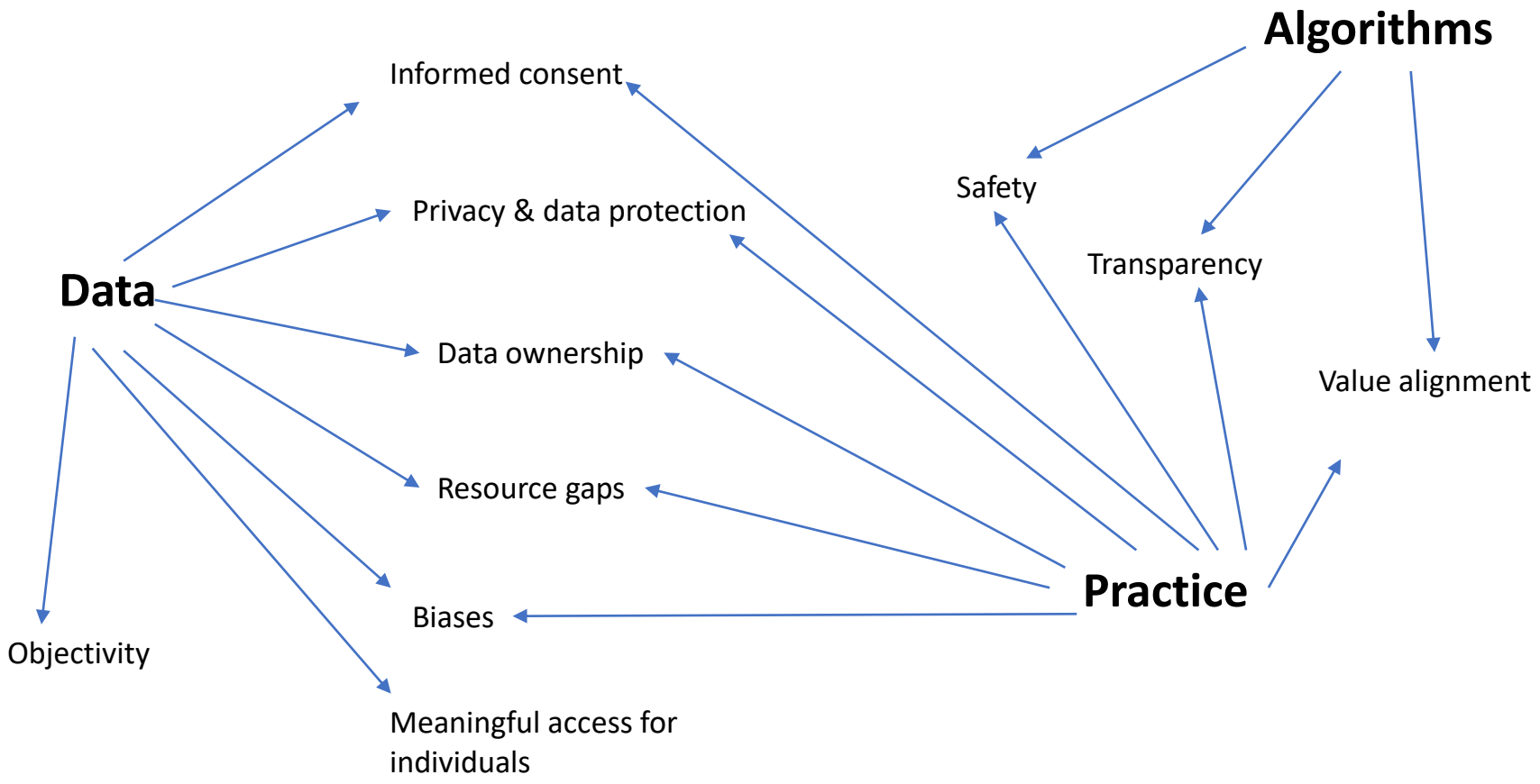
University College, Cork, Ireland

Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement

J. Raymond Geis, MD • Adrian P. Brady, MB, FFRRCSI • Carol C. Wu, MD • Jack Spencer, PhD • Erik Ranshaert, MD, PhD • Jacob L. Jaremko, MD, PhD • Steve G. Langer, PhD • Andrea Borondy Kitts, MS, MPH • Judy Birch, BEd • William F. Shields, JD, LLM • Robert van den Hoven van Genderen, PhD, MSc, LLM • Elmar Kotter, MSc, MD, MBA • Judy Wawira Gichoya, MBChB, MS • Tessa S. Cook, MD, PhD • Matthew B. Morgan, MD, MS • An Tang, MD, MSc • Nabile M. Safdar, MD, MPH • Marc Kohli, MD

- **ACR, ESR, RSNA, CAR, SIIM, EuSoMII, AAPM**
- Joint publication October 2019, in
  - **Journal of the American College of Radiology (JACR)**
  - **Insights into Imaging** 2019;10:101. https://doi.org/10.1186/s13244-019-0785-8
  - **Radiology** 2019; 00:1–5 • https://doi.org/10.1148/radiol.2019191586
  - **Canadian Association of Radiologists Journal**




Algorithms
Informed consent
Safety
Privacy & data protection
Transparency
Data
Data ownership
Value alignment
Resource gaps
Practice
Biases
Objectivity
Meaningful access for individuals

Geoffrey Hinton, Toronto 2016

"If you work as a radiologist, you're like the coyote that's already over the edge of the cliff… People should stop training radiologists now. It's just completely obvious that within 5 years, deep learning is going to be better than radiologists."

‘Godfather of AI’ Geoffrey Hinton quits Google and warns over dangers of misinformation

The neural network pioneer says dangers of chatbots were ‘quite scary’ and warns they could be exploited by ‘bad actors’

Josh Taylor *and* **Alex Hern**

Tue 2 May 2023 12.23 BST

The Guardian
Newspaper of the year

The Guardian
Newspaper of the year

- “Hinton … said he quit to speak freely about the dangers of AI, and in part regrets his contribution to the field….”
- “Some of the dangers of AI chatbots were ‘quite scary’, he told the BBC, warning they could become more intelligent than humans, and could be exploited by bad actors”
- “Hinton’s concern … is …that … people will not be able to discern what is true any more with AI”
- “Hinton was also concerned that AI will eventually replace jobs like paralegals, personal assistants and other ‘drudge work’”

Elon Musk joins call for pause in creation of giant AI 'digital minds'

March 29, 2023

More than 1,000 artificial intelligence experts urge delay until world can be confident 'effects will be positive and risks manageable'

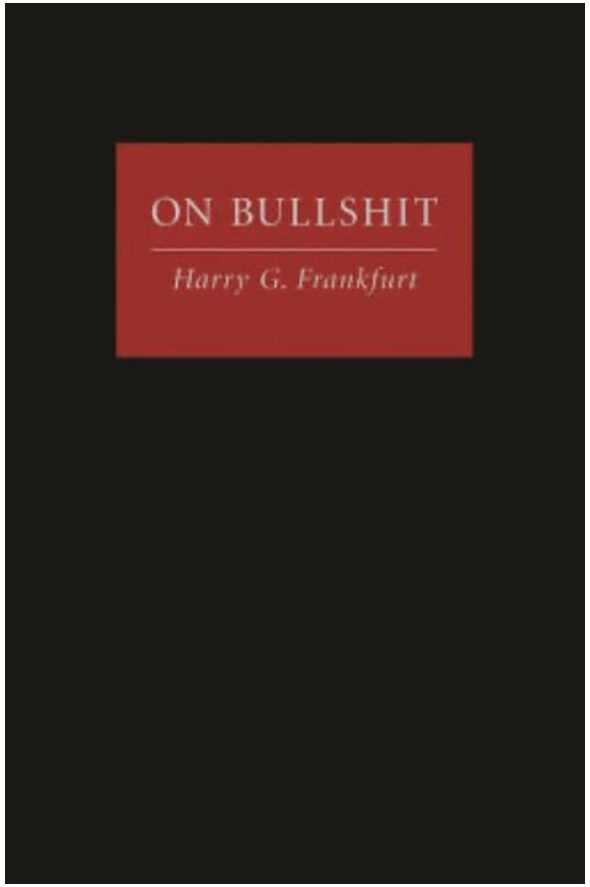
The Guardian
Newspaper of the year

A letter co-signed by Elon Musk and thousands of others demanding a pause in artificial intelligence research has created a firestorm, after the researchers cited in the letter condemned its use of their work, some signatories were revealed to be fake, and others backed out on their support.

April 1, 2023

Ethical dilemma No. 1 – Truth or fiction?

- “It is just this lack of connection to a concern with truth – this indifference to how things really are – that I regard as of the essence of bullshit.”
- “For Chat GPT, values do not exist. It produces words, but it has no way of caring about their meaning in the world. Machines do not have intentions. People do.” (Fintan O’Toole, Irish Times 13/5/23)



Raritan Quarterly Review, 1986
Princeton University Press 2005


GENERATED BY A.I.


GENERATED BY A.I.


GENERATED BY A.I.


GENERATED BY A.I.

CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning

Yisroel Mirsky[1], Tom Mahler[1], Ilan Shelef[2], and Yuval Elovici[1]
[1]Department of Information Systems Engineering, Ben-Gurion University, Israel
[2]Soroka University Medical Center, Beer-Sheva, Israel
yisroel@post.bgu.ac.il, mahlert@post.bgu.ac.il, shelef@bgu.ac.il, and elovici@bgu.ac.il

Published in the 28th USENIX Security Symposium (USENIX Security 2019)

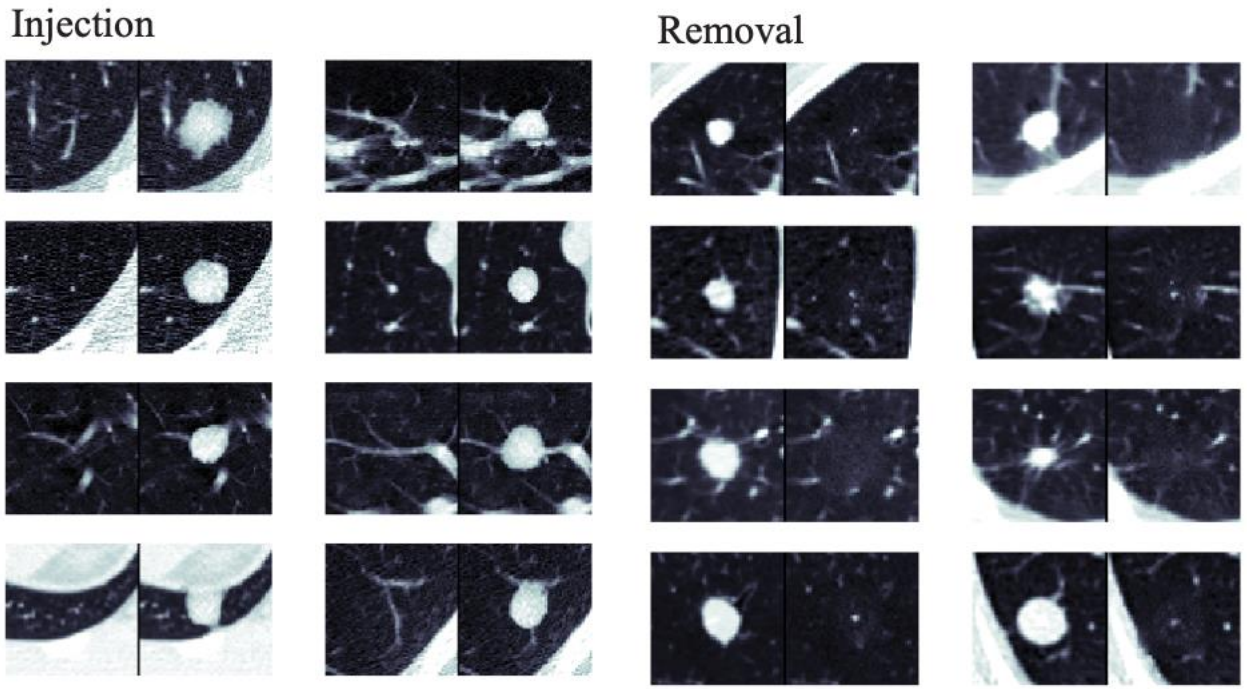


Figure 9: Sample injections (left) and removals (right). For each image, the left side is before tampering and the right side is after. Note that only the middle 2D slice is shown.

Ethical dilemma No. 2 - Generative AI and research/publishing

1. AI being researched to eliminate key roles of radiologists

Radiology

GPT-4 for Automated Determination of Radiological Study and Protocol based on Radiology Request Forms: A Feasibility Study

Manuscript Type: Original Research (Research Letter)

Authors: Roman Johannes Gertz[a], Alexander Christian Bunck[a], Simon Lennartz[a], Thomas Dratsch[a], Andra-Iza Iuga[a], David Maintz[a], Jonathan Kottlors[a]

Affiliations:

[a]Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

"We aimed to evaluate the performance of GPT-4 at selecting correct imaging studies and protocols based on medical history and corresponding clinical questions extracted from RRFs"

July 2023

Towards Generalist Biomedical AI

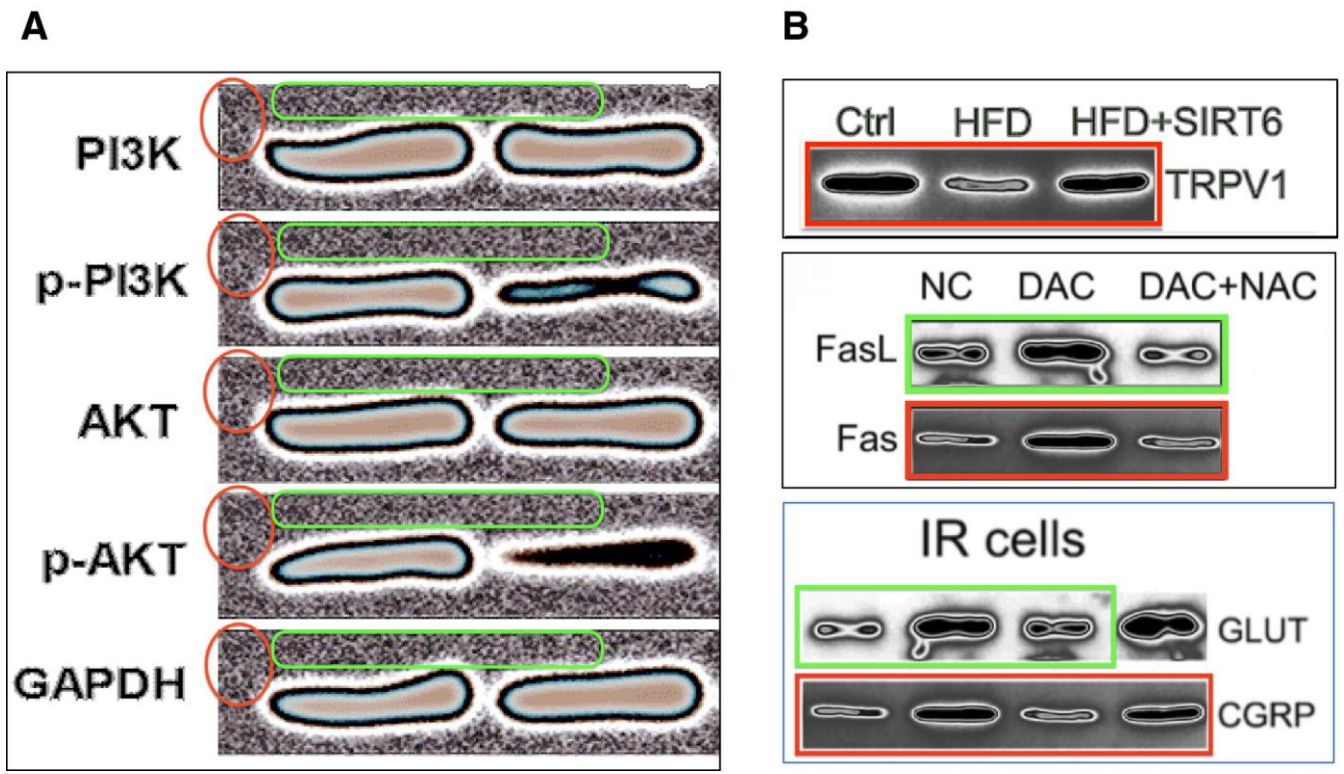
Tao Tu*, ‡, 1, Shekoofeh Azizi*, ‡, 2,
Danny Driess[2], Mike Schaekermann[1], Mohamed Amin[1], Pi-Chuan Chang[1], Andrew Carroll[1],
Chuck Lau[1], Ryutaro Tanno[2], Ira Ktena[2], Basil Mustafa[2], Aakanksha Chowdhery[2], Yun Liu[1],
Simon Kornblith[2], David Fleet[2], Philip Mansfield[1], Sushant Prakash[1], Renee Wong[1], Sunny Virmani[1],
Christopher Semturs[1], S Sara Mahdavi[2], Bradley Green[1], Ewa Dominowska[1], Blaise Aguera y Arcas[1],
Joelle Barral[2], Dale Webster[1], Greg S. Corrado[1], Yossi Matias[1], Karan Singhal[1], Pete Florence[2],
Alan Karthikesalingam†, ‡,1 and Vivek Natarajan†, ‡,1

[1]Google Research, [2]Google DeepMind

arXiv:2307.14334v1 [cs.CL] 26 Jul 2023

"…we conduct a radiologist evaluation of model-generated (and human) chest X-ray reports and observe encouraging performance across model scales. In a side-by-side ranking of 246 retrospective chest X-rays, clinicians express a pairwise preference for Med-PaLM M reports over those produced by radiologists in up to 40.50% of cases, suggesting potential clinical utility."

2. AI being used to fake data



17 biochemistry-related studies

3 journals, several months

Same background or repeated use of same images (with different protein identification)

? Paper-mill

Digital magic, or the dark arts of the 21st century—how can journals and peer reviewers detect manuscripts and publications from paper mills?

FEBS Letters, Volume: 594, Issue: 4, Pages: 583-589, First published: 17 February 2020, DOI: (10.1002/1873-3468.13747)

You Can Just Put Up A Poster At ICML

and Nobody Will Stop You

Jimmy Neuron, Chet G. P. Tee IV, Ada Grahd

FedEx is All You Need

Background

A perennial problem in machine learning research has been how to most efficiently have a poster at ICML. In this work, we show that one can bypass OpenReview via a simple "FedEx trick," similar to yet entirely different from the kernel trick in machine learning. This represents a substantial improvement over the prior SOTA [1].

Research Question: Can you just put up a poster at ICML?
Answer: yes.

Main Contribution

We introduce a novel *FedEx attack* on the conference review system:

1) Hastily design a poster in a hotel room.
2) Print it and bring to Thursday poster session.
3) Surreptitiously mount it on an unused board.

This attack opens new avenues of research into defenses, including embodied human agents fine-tuned to detect and remove spurious content.

Future work

For simplicity, this work treats (and consists of) a poster with only one layer. Future work will focus on deep posters with many layers.

[1] You Can Just Put Up A Poster At AISTATS. JN, CT, AG. *AISTATS 2023.*

Theory

Theorem 1 (Free-Lunch Theorem for Posters)

The Free-Lunch theorem demonstrates that one can secure a poster at ICML without any discernible work or innovative contribution. Formally, let $\wp$ denote this poster, and let $\P$ denote the set of posters presented at ICML. This theorem then states $\boxed{\wp \in \P}$.

The following is a short sketch of the proof. Let $\mathscr{C}$ denote a given review committee, and let $\mathscr{P}$ and $\mathscr{R}$ denote other symbols. Now then, for a given $\wp$, we have:

$$\wp \sim \mathscr{R} \cap (*\mathscr{C})$$

$$\Rightarrow \forall \wp^*, \langle \wp, \mathscr{P} - \mathscr{R} | \wp, \mathscr{P} - \mathscr{R} \rangle \mapsto \wp^*$$

$$\Rightarrow \sum_{i=0}^{|\P|} [\wp_i \cdot (*\mathscr{C})] + \frac{\langle \wp_i, \mathscr{P} - \mathscr{R} | \wp_i, \mathscr{P} - \mathscr{R} \rangle}{|\P|} < |\P| \qquad \text{(FedEx trick)}$$

By "$\rtimes$"-ing the previous equation with $(*\mathscr{C})$ once again, we have:

$$\frac{\langle \wp^*, \mathscr{P} - \mathscr{R} | \wp^*, \mathscr{P} - \mathscr{R} \rangle}{|\P|} \sim \mathscr{C}$$

$$\Rightarrow \boxed{\wp \in \P}$$

Your reading this demonstrates this result exactly matches experiment.

As is customary, we also include an additional theorem unrelated to our main contribution:

Theorem 2 (Universal Approximation Theorem for Posters)

Let $\mathcal{P}_w$ denote all subsets of $\{0,1\}^{\mathrm{N}}$ expressible using posters of width w. Then the following holds for every $U \in \{0,1\}^{\mathrm{N}}$:

$$\lim_{w \to \infty} \inf_{P \in \mathcal{P}_w} d_{\mathrm{H}}(P, U) = 0$$

Where d_{H} denotes the Hamming distance.

Empirics

All experiments were run on a Game Boy Advance™ running Arch Linux, which allows our methods to easily extend to edge computing use cases. We determined the poster width using a hyperparameter grid search with a grid size of 1 (as is standard). Results are shown in Figure 2 below.

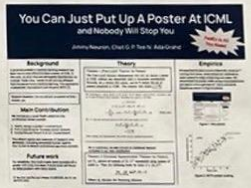


Figure 1: this poster.

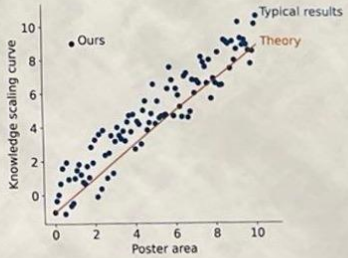


Figure 2: SOTA poster scaling.

3. AI being used to write text

- Benefits
  - Better writing quality, structure
  - Language editing
- AI-like tools used by journals
    - Checking for duplicate, plagiarized or fake manuscripts
    - Assigning reviewers, etc.
- Using generative AI breaks confidentiality of data (sharing with everybody)

4. AI being used to write peer reviews

- Paper preprint from own group
- ChatGPT wrote review
  - Good summary
  - Some positive subjective comments
    - “The writing style is clear and concise”
  - “When asked to suggest more specific improvements, it fails and starts … hallucinating”
    - Specific-looking general comments with no bearing on text
  - When asked for additional references
    - Real-sounding but non-existent articles

Donker T, The dangers of using large language models for peer-review. Lancet Infectious Diseases (2023). DOI: 10.1016/S1473-3099(23)00290-6

International Committee of Medical Journal Editors (ICMJE) recommendations May 2023

- AI or AI-assisted technologies do not qualify as authors and must not be listed
- At submission, authors must disclose whether they used AI in preparation of manuscript
- All authors are responsible for any submitted material that includes AI-assisted technologies
- Authors should be able to assert that there is no plagiarism in their paper, including in text and images produced by AI
- Reviewers should not upload manuscript to software or other AI technologies where confidentiality cannot be assured
- Use of AI to assist in peer review would involve a breach of confidentiality

General Omar Bradley, Nov. 10, 1948, Boston Chamber of Commerce

- “The world has achieved brilliance without wisdom, power without conscience. Ours is a world of nuclear giants and ethical infants.”