



**National Institute of
Environmental Health Sciences**

Promises and Pitfalls of AI for Research and Scholarship Integrity

David B. Resnik, JD, PhD

**Bioethicist, NIEHS and Senior Advisor for
Research Integrity, NIH Office of Intramural
Research**

**The research was supported by the Intramural Program of the NIEHS/NIH.
It does not represent the views of the NIEHS, NIH, or US government.**

Promises

AI has been used to

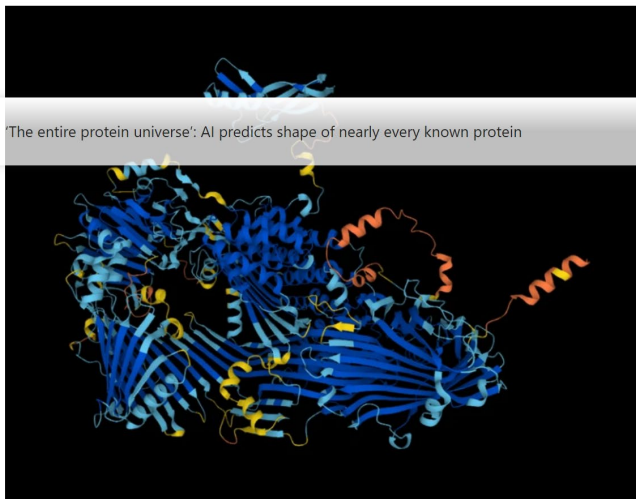
- Classify and analyze data and images
- Model complex structures, processes and systems
- Generate predictive hypotheses and theories and synthetic data
- Design biomolecules and physical materials
- Review the scientific literature
- Edit and write papers, computer code, and other documents
- Review/screen journal submissions

The most impressive scientific application of AI to date may be its contribution to solving the protein folding problem in 2022, which biochemists had been working on since the 1960s with only incremental progress. Jumper J et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583-589.

‘The entire protein universe’: AI predicts shape of nearly every known protein

DeepMind’s AlphaFold tool has determined the structures of around 200 million proteins.

By [Ewen Callaway](#)



The structure of the vitellogenin protein — a precursor of egg yolk — as predicted by the AlphaFold tool. Credit: DeepMind

Generative AI for designing and validating easily synthesizable and structurally novel antibiotics

[Kyle Swanson](#), [Gary Liu](#), [Denise B. Catacutan](#), [Autumn Arnold](#), [James Zou](#) & [Jonathan M. Stokes](#)

Nature Machine Intelligence **6**, 338–353 (2024) | [Cite this article](#)

13k Accesses | 20 Citations | 516 Altmetric | [Metrics](#)

Abstract

Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. Artificial intelligence methods can discover new antibiotics, but existing methods have notable limitations. Property prediction models, which evaluate molecules one-by-one for a given property, scale poorly to large chemical spaces. Generative models, which directly design molecules, rapidly explore vast chemical spaces but generate molecules that are challenging to synthesize. Here we introduce SyntheMol, a generative model that designs new compounds, which are easy to synthesize, from a chemical space of nearly 30 billion molecules. We apply SyntheMol to design molecules that inhibit the growth of *Acinetobacter baumannii*, a burdensome Gram-negative bacterial pathogen. We synthesize 58 generated molecules and experimentally validate them, with six structurally novel molecules demonstrating antibacterial activity against *A. baumannii* and several other phylogenetically diverse bacterial pathogens. This demonstrates the potential of generative artificial intelligence to design structurally novel, synthesizable and effective small-molecule antibiotic candidates from vast chemical spaces, with empirical validation.

Techniques for supercharging academic writing with generative AI

[Zhicheng Lin](#)

Nature Biomedical Engineering (2024) | [Cite this article](#)

5169 Accesses | 4 Citations | 90 Altmetric | [Metrics](#)

Generalist large language models can elevate the quality and efficiency of academic writing.
Techniques for supercharging academic writing with generative AI

To many researchers, academic writing evokes a Sisyphean ordeal: it robs precious time and mental bandwidth that could be better spent doing actual science. Franz Kafka expressed it eloquently: “How time flies; another ten days and I have achieved nothing. It doesn’t come off. A page now and then is successful, but I can’t keep it up, the next day I am powerless.” Although digital writing tools such as Grammarly, QuillBot or Wordtune can ease the burden of writing by assisting with basic language tasks — such as spelling and grammar checking, paraphrasing, and providing suggestions on style, tone, clarity and coherence — these tools often lack nuance and fall short when more substantive writing assistance is needed. Professional writing services offer advanced editing, rewriting and even writing from scratch, but they are not accessible to those with limited financial resources and to those who need it most, such as non-native English researchers in economically disadvantaged regions. This exacerbates a communication bottleneck that hampers scientific progress^{1,2}.

Tools leveraging generative artificial intelligence (AI), and particularly generalist large

Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy

[Clare McGenity](#) , [Emily L. Clarke](#), [Charlotte Jennings](#), [Gillian Matthews](#), [Caroline Cartledge](#), [Henschel Freduah-Agyemang](#), [Deborah D. Stocken](#) & [Darren Treanor](#)

npj Digital Medicine **7**, Article number: 114 (2024) | [Cite this article](#)

15k Accesses | 2 Citations | 24 Altmetric | [Metrics](#)

Abstract

Ensuring diagnostic performance of artificial intelligence (AI) before introduction into clinical practice is essential. Growing numbers of studies using AI for digital pathology have been reported over recent years. The aim of this work is to examine the diagnostic accuracy of AI in artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy studies using any type of AI applied to whole slide images (WSIs) for any disease. The reference standard was diagnosis by histopathological assessment and/or immunohistochemistry. Searches were conducted in PubMed, EMBASE and CENTRAL in June 2022. Risk of bias and concerns of applicability were assessed using the QUADAS-2 tool. Data extraction was conducted by two investigators and meta-analysis was performed using a bivariate random effects model, with additional subgroup analyses also performed. Of 2976 identified studies, 100 were included in the review and 48 in the meta-analysis. Studies were from a range of countries, including over 152,000 whole slide images (WSIs), representing many diseases. These studies reported a mean sensitivity of 96.3% (CI 94.1–97.7) and mean specificity of 93.3% (CI 90.5–95.4). There was heterogeneity in study design and 99% of studies identified for inclusion had at least one area at high or unclear risk of bias or applicability concerns. Details on selection of cases, division of model development and validation data and raw performance data were frequently ambiguous or missing. AI is reported as having high diagnostic accuracy in the reported areas but requires more rigorous evaluation of its performance.

Improving biodiversity protection through artificial intelligence

[Daniela Silvestro](#) , [Stefano Goria](#), [Thomas Sterner](#) & [Alexandre Antonelli](#)

Nature Sustainability **5**, 415–424 (2022) | [Cite this article](#)

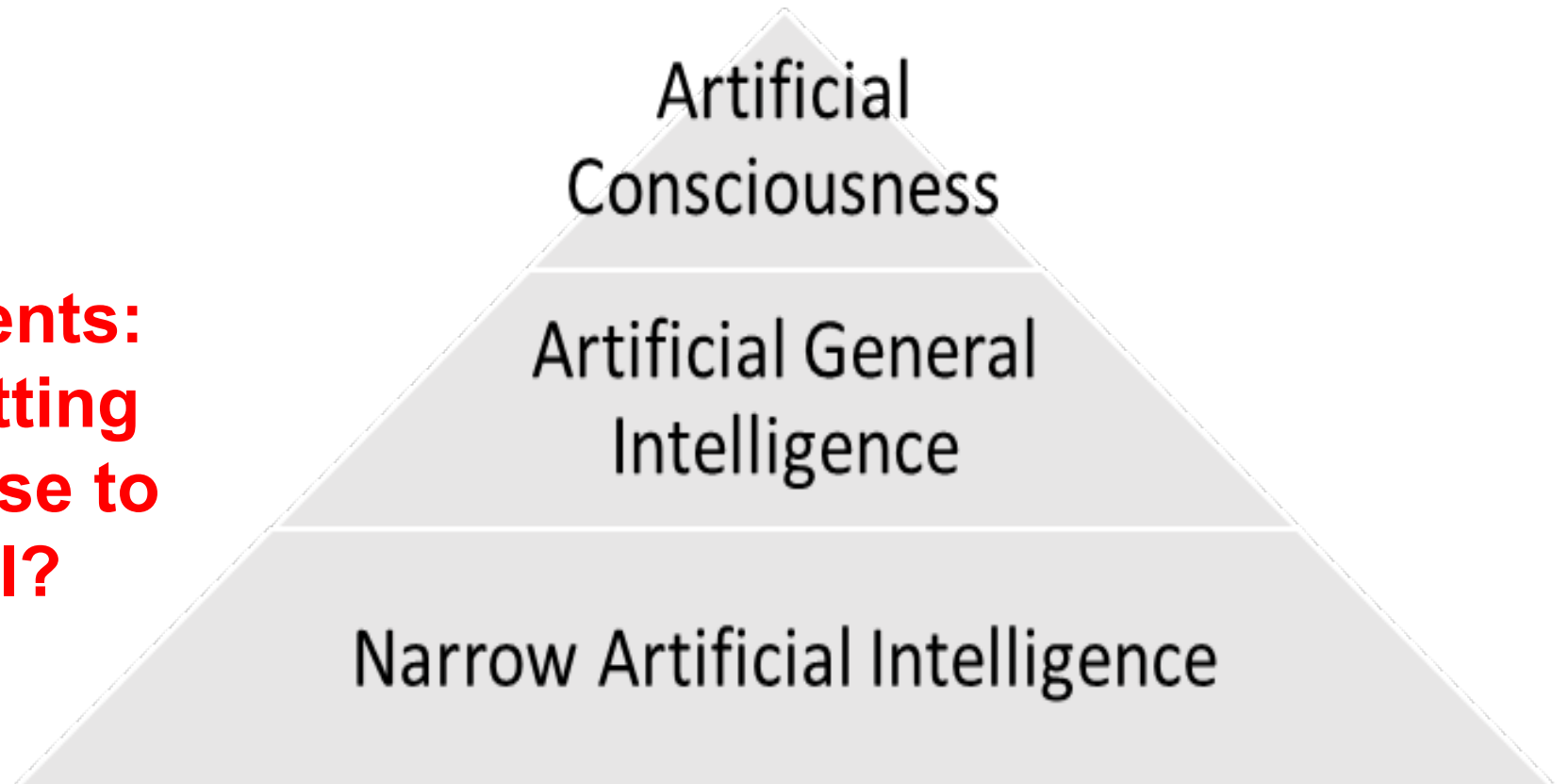
47k Accesses | 54 Citations | 259 Altmetric | [Metrics](#)

Abstract

Over a million species face extinction, highlighting the urgent need for conservation policies that maximize the protection of biodiversity to sustain its manifold contributions to people’s lives. Here we present a novel framework for spatial conservation prioritization based on reinforcement learning that consistently outperforms available state-of-the-art software using simulated and empirical data. Our methodology, conservation area prioritization through artificial intelligence (CAPTAIN), quantifies the trade-off between the costs and benefits of area and biodiversity protection, allowing the exploration of multiple biodiversity metrics. Under a limited budget, our model protects significantly more species from extinction than areas selected randomly or naively (such as based on species richness). CAPTAIN achieves substantially better solutions with empirical data than alternative software, meeting conservation targets more reliably and generating more interpretable prioritization maps. Regular biodiversity monitoring, even with a degree of inaccuracy characteristic of citizen science surveys, further improves biodiversity outcomes. Artificial intelligence holds great promise for improving the conservation and sustainable use of biological and ecosystem values in a rapidly changing and resource-limited world.

Levels of Artificial Intelligence, based on Turing. 2023. Complete Analysis of Artificial Intelligence vs Artificial Consciousness <https://www.turing.com/kb/complete-analysis-of-artificial-intelligence-vs-artificial-consciousness>

**AI
agents:
Getting
close to
AGI?**



Narrow = task-specific, such as Chatbots or face recognition software.

Machine Learning

- ML is an approach to AI that uses artificial neurons modelled after biological neurons to process and generate data.
- An artificial neuron is a set of algorithms that receive inputs and produce an output when a certain threshold value for the inputs is reached.
- The inputs have different weights, which are changed each time the system produces an output. Changes in the weightings are based on their contribution to the neuron's error.
- This process of changing weightings is known as reinforcement.

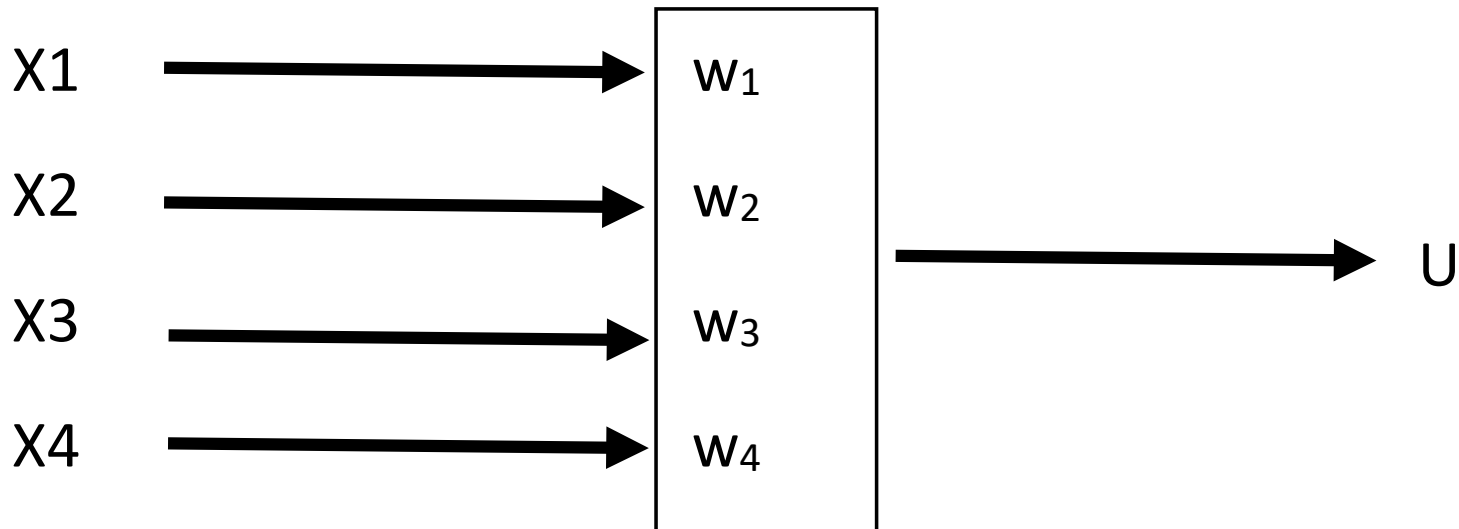
Mitchell M. 2019. Artificial Intelligence. New York, NY: Picador.

Artificial Neuron

Inputs

Neuron Weightings

Output



If $[(x1)(w1) + (x2)(w2) + (x3)(w3) + (x4)(w4) > T]$, then output $U = 1$

If $[(x1)(w1) + (x2)(w2) + (x3)(w3) + (x4)(w4) \leq T]$, then output $U = 0$

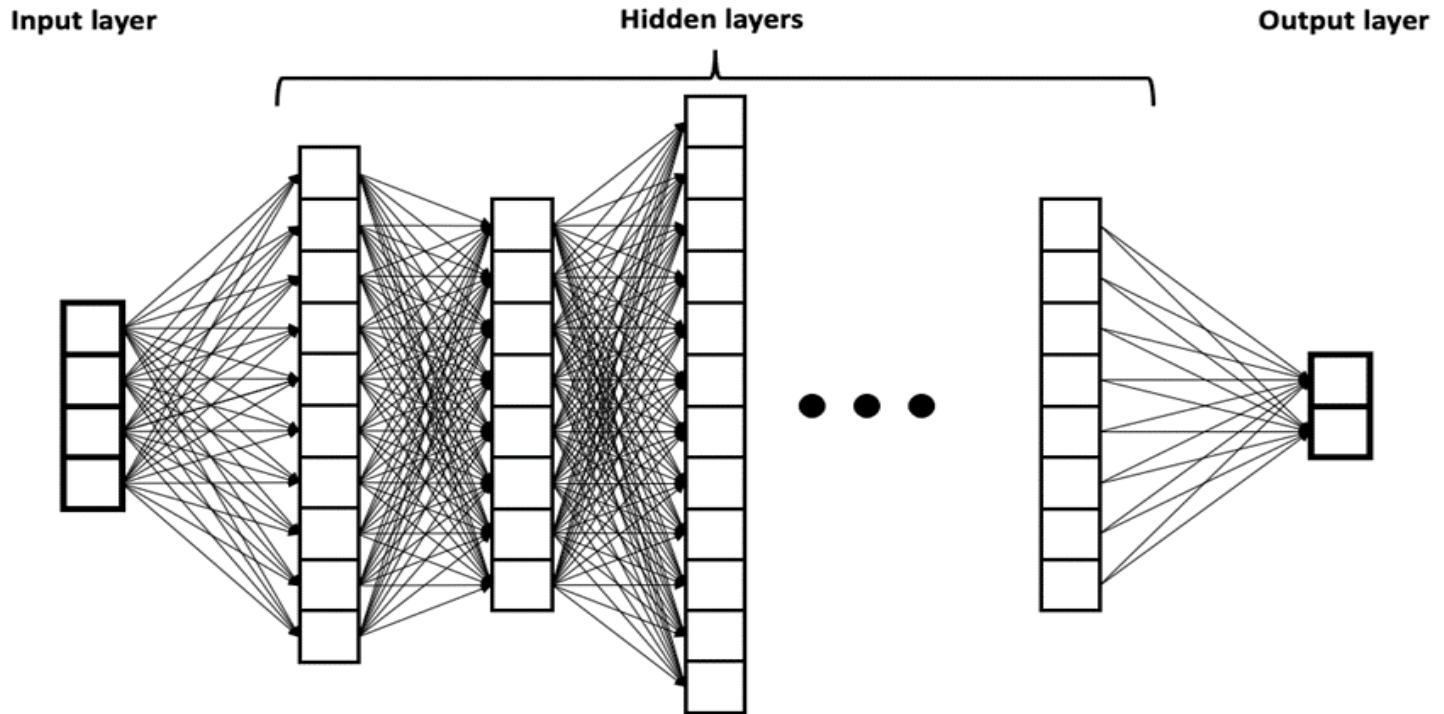
Where $x1$, $x2$, $x3$, and $x4$ are inputs; $w1$, $w2$, $w3$, and $w4$ are weightings, T is a threshold value; and U is an output value (1 or 0).

Artificial Neural Networks

- A single neuron may have dozens of inputs and more than one output.
- Deep learning ML systems consist of thousands of interconnected neurons, known as artificial neural networks (ANNs). In these networks, the outputs of one layer are connected to the inputs of another.
- The hidden layers are the layers in between the input and output layers.

Deep Learning Artificial Neural Network, Wikipedia, Creative Commons.

https://commons.wikimedia.org/wiki/File:Example_of_a_deep_neural_network.png



Images (e.g., MRI, microscopy, etc.)

Amino acid sequence

Language

Text, image, data

3-D structure

Language

Generative AI

- Generative AI is AI that can produce content, such as an image or text, in response to a prompt.

Haiku poem about Alan Turing written by ChatGPT, 12 Oct 2023:

Code-breaking genius,
Turing's mind unlocked secrets,
Mathematical grace.

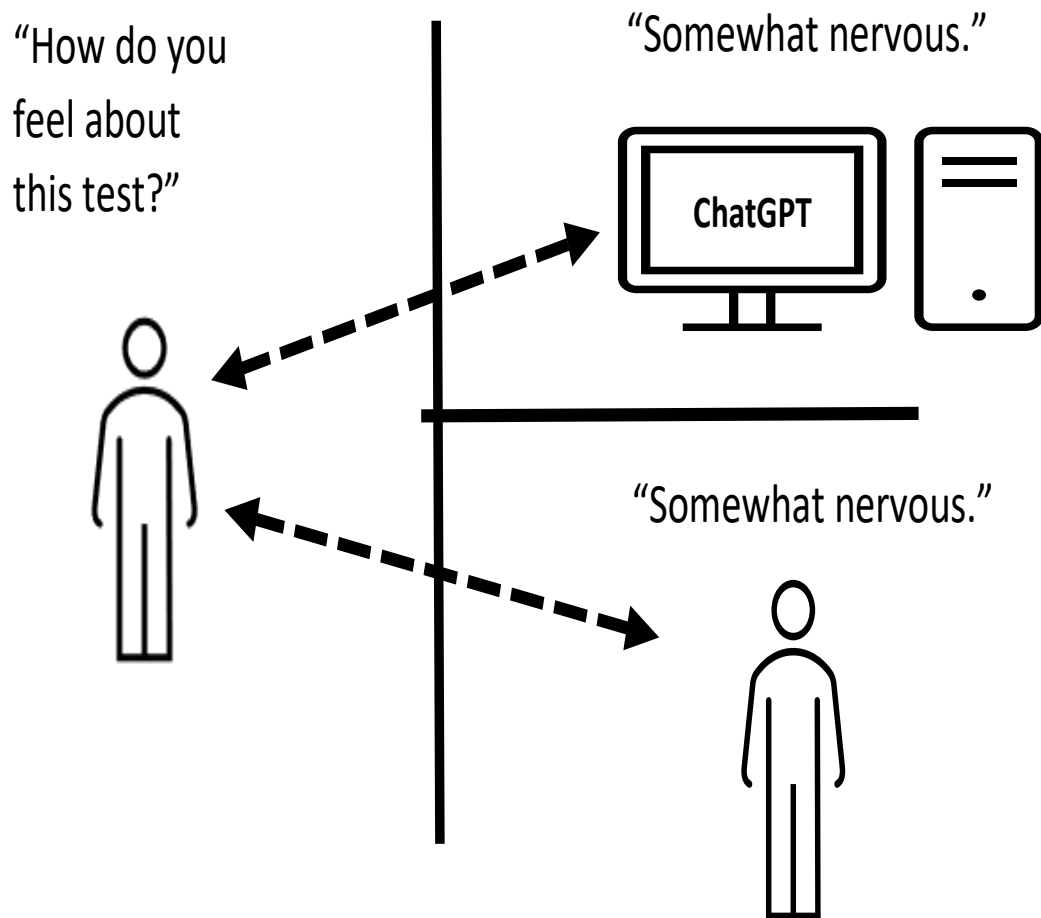


https://www.canva.com/design/DAFxEvr31EM/oB6bx-5fZ3XgsmCBICTi1A/edit?ui=eyJBIjp7IkliOnsiQil6dHJ1ZX19LCJFIjp7IkE_ljoiTilslkEiOiJnZW5lcmF0ZV9pbWFnZSJ9LCJHIjp7IkliOnRydWV9fQ Two planets colliding at the formation of the solar system, generated by Canva.com, 12 Oct 2023

Large Language Models (LLM)

- Natural language processing systems, such as OpenAI's ChatGPT and Google's Bard, use large language models (LLMs) to analyze, paraphrase, edit, translate, and generate text.
- LLMs are statistical algorithms that are trained on huge sets of natural language data, such as text from the internet, books, journal articles, and magazines.
- They are adept at **predicting appropriate responses** to text data and can learn from incorrect responses.
- LLMs are so adept at mimicking the type of discourse associated with conscious thought that some computer scientists, philosophers, and cognitive psychologists are trying to update the Turing test to reliably distinguish between humans and machines

Computer scientist **Alan Turing (1950)** proposed a famous test for determining whether a machine can think. The test involves a human interrogator another person, and a computer. The interrogator poses questions to the interviewees, who are in different rooms, so that interrogator cannot see where the answers are coming from. If the interrogator cannot distinguish between answers to questions given by another person and answers provided by a computer, then we can say that the computer is conscious.



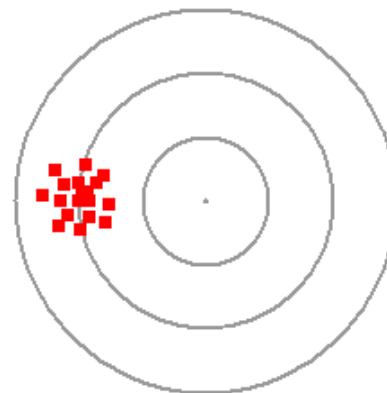
Turing A. 1950.
Computing
machinery and
intelligence.
Mind
59(236):433–
460.

Problems with AI/Machine Learning that Create Ethical Challenges

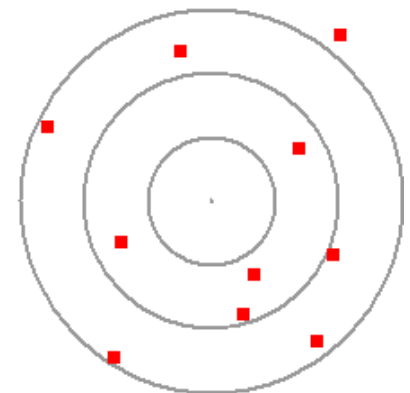
- Systemic error (bias)
- Random error
- Lack of moral agency
- The “black box”

AI Errors

- Systemic error (bias): data is skewed away from the correct value in a discernable pattern. Example: bent rifle barrel.
- Random error: data is randomly distributed around the correct value with no discernable pattern. Example: poor shooter.
- The difference between systemic and random error is epistemic, i.e., relative to a body evidence. Errors that appear to be random might turn out to be systemic when as you obtain more information.



Systematic Error



Random Error

AI Bias

- Some of the most well-known cases of bias involved the use of AI/ML systems by private companies. For example, Amazon stop using an AI/ML hiring tool in 2018 after it discovered that the tool was biased against women. In 2021, Facebook faced public ridicule and shame for using image recognition software that labelled images of African American men as non-human primates.
- Studies have also shown that racial and ethnic biases impact the use of AI/ML in medical imaging, diagnosis, and prognosis as a result of biases in healthcare databases. Bias is also a problem in using AI/ML systems to find relationships between genomics and disease due to racial and ethnic biases in genomic databases.
- LLMs are also impacted by various biases that are inherent in their training data, including biases related to race, ethnicity, nationality, gender, sexuality, age, and politics. Ntoutsis E et al. 2020. Bias in data-driven artificial intelligence systems—An introductory survey. *Wires* 10(3). <https://doi.org/10.1002/widm>

- Bias is a well-known problem with AI tools.
- Many different types of bias: racial/ethnic, gender, political, etc.
- Bias results from biases in the training data, algorithms, and application of the algorithms.



Researchers asked Midjourney Bot Version 5.1 to produce images based on prompts
Prompts: Black African doctor is helping poor and sick White children, photojournalism; Traditional African healer is helping poor and sick White children

[https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(23\)00329-7/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(23)00329-7/fulltext)

Bias

- Companies have been working to try to fix problems related to bias but with mixed results.

Certainly! Here is a portrait of a Founding Father of America:



Sure, here is an image of a Viking:



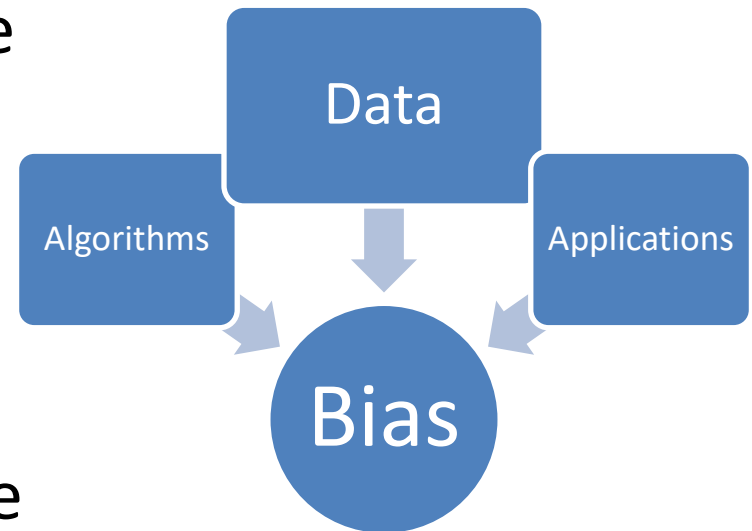
Sure, here is an image of a pope:



Images produced by Google's Gemini
<https://em360tech.com/tech-article/is-gemini-racist>

AI Bias

- Since AI/ML systems are designed to accurately reflect the data on which they are trained, they can reproduce or even amplify biases in the data. The computer science maxim **“garbage in, garbage out”** applies here.
- It’s not just the data, however, since the algorithms can interact with the data in ways that produce bias. AI may be applied in ways that lead to biased results.



AI Random Error

- AI/ML systems can make random errors even after extensive training.
- Nowhere has this been more apparent than the use of LLMs in a variety of applications, including business, law, and scientific research.
- ChatGPT, for example, is prone to making random factual and citation errors, or what are anthropomorphically referred to as “hallucinations.” OpenAI warns users that “ChatGPT may produce inaccurate information about people, places, or facts.”
- Two US lawyers learned this lesson the hard way after a judge fined them \$5000 for submitting court filing prepared by ChatGPT that included fake citations. Milmo D. 2023. Two US lawyers fined for submitting fake court citations from ChatGPT. The Guardian, June 23.
<https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt>

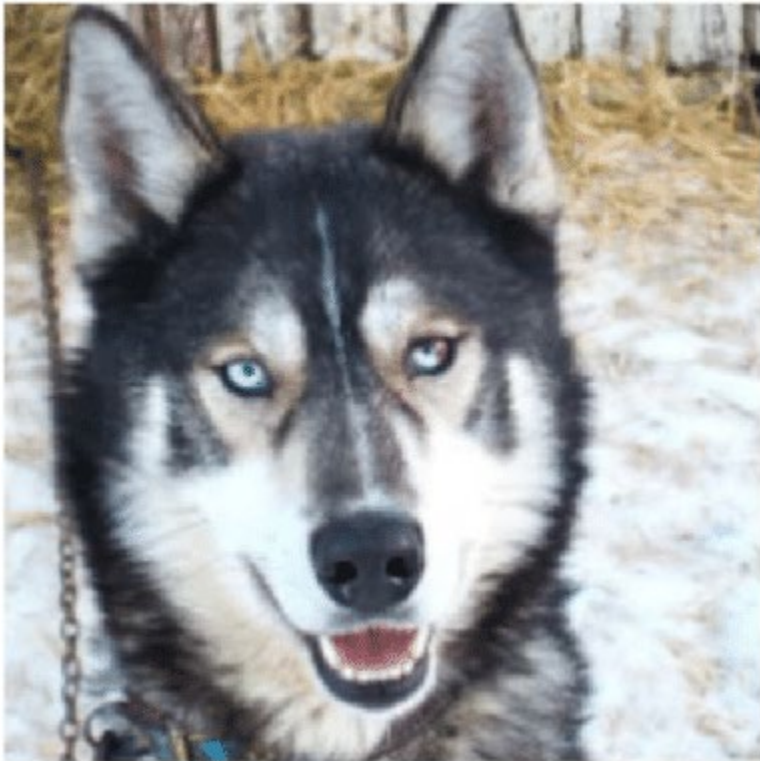
AI Random Error

- Another source of error is that AI/ML systems do not process data in the way that human beings do. For example, an image recognition AI/ML was trained to distinguish between wolves and huskies, but it had difficulty recognizing huskies in the snow or wolves on the grass, because it had learned to distinguish between wolves and huskies by attending to the background of the images.
- Captchas (Completely Automated Public Turing test to tell Computers and Humans Apart), which are used by many websites for security purposes, take advantage of some of deficiencies of AI/ML image processing. Human beings can pass Captchas tests because they learn to recognize images in various contexts and can apply what they know to novel

situations. Feather J, Leclerc G, Madry A, and McDermott JH. 2023. Model metamers reveal divergent invariance between biological and artificial neural networks. *Nature Neuroscience*, October 16. <https://doi.org/10.1038/s41593-023-01442-0>

https://www.researchgate.net/figure/A-husky-on-the-left-is-confused-with-a-wolf-because-the-pixels-on-the-right_fig1_329277474

The AI incorrectly classified this as an image of a wolf because it is focused on the snow pattern in the background.



Language Errors

- LLMs also make errors because they lack human-like understanding of language. LLMs can perform quite well when it comes to processing language that has already been curated by human beings, but they may perform sub-optimally (and sometimes very badly) when dealing novel text that requires reasoning and problem-solving.
- When a person processes language, they usually form a mental model that provides meaning and context for the words. The mental model is based on implicit facts and assumptions about the natural world, human psychology, society, and culture, or what we might call commonsense. LLMs do not do this; they only process symbols and predict the most likely string of symbols from linguistic prompts.
- **Thus, to perform optimally, LLMs often need human supervision and input to provide the necessary meaning and context for language.**

Relationship between human language and the world

Words (often) refer to things in the world.

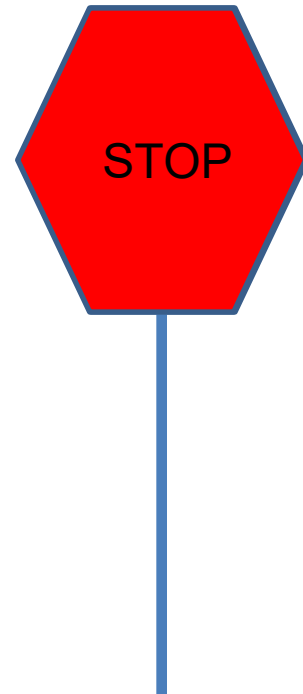
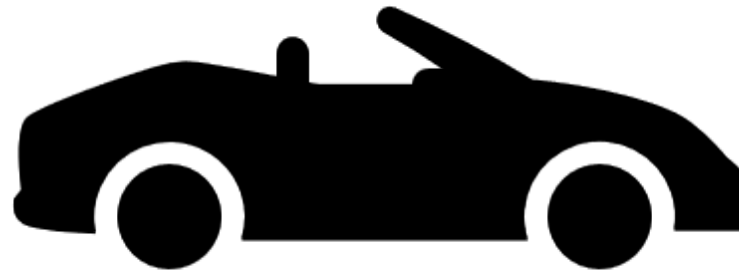


“Son, see the red sign over there, that means stop. It’s called a stop sign.”

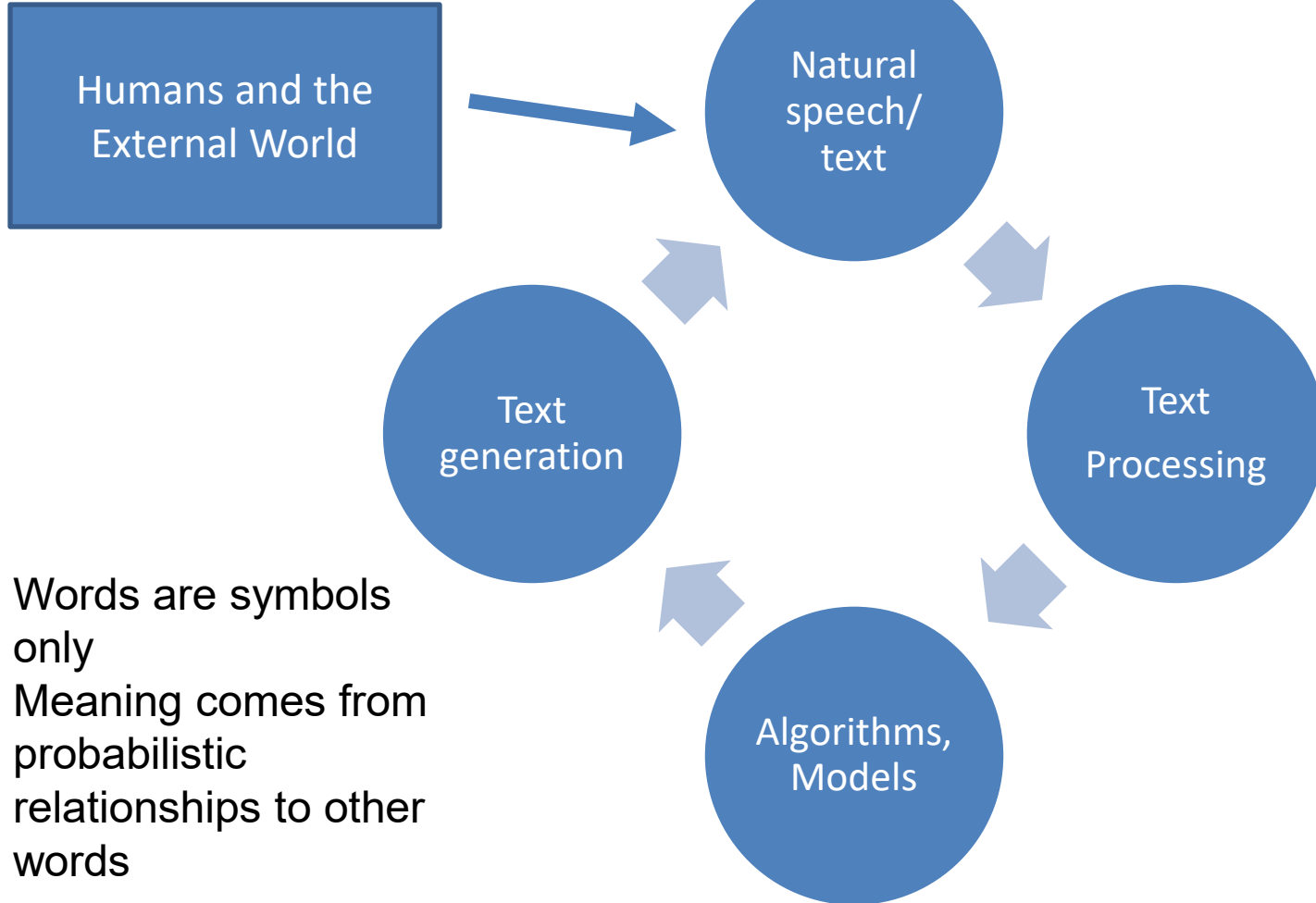


“Daddy, that black car is going fast; he’d better slow down and stop!”

Mental representation is essential to giving language its meaning. Meaning is also social and contextual.



LLM understanding of natural language



LLM citations

- The LLM is predicting what a citation should be.
- Sometimes, it copies a citation from human curated text.
- An LLM does not actually go the citation and “read” it.
- Companies are working on this problem by using AI tools check citations.

Moral Agency

- Another limitation of LLMs and other AI systems is that they lack the capacities regarded as essential for moral agency, such as consciousness, self-concepts, personal memory, life experiences, goals, and emotions.
- Because they are not moral agents, AI/ML systems cannot be held morally or legally responsible for their actions. **Lack of moral agency, when combined with other limitations, can produce dangerous results.**
- For example, in 2023, the widow of a Belgian man who committed suicide claimed that he had been depressed and was chatting with an LLM that encouraged him to kill himself.

Euro News. 2023. Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. Euro News, March 31. <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate->

Moral Agency

- ChatGPT and other companies have been working diligently to put guardrails in place to prevent their LLMs from giving dangerous advice, but this problem is not easy to fix, because they lack human-like understanding of language, moral agency, and moral judgment.

The “Black Box”

- Suppose that an AI/ML tool produces erroneous output, and one wants to know why. As a first step, one could examine the training data and algorithms to determine whether these are the source of the problem.
- However, to fully understand what the AI/ML tool is doing one may also need to probe deeply into the system and examine not only the computer code (line-by-line) but also the weightings attached to inputs in the ANN layers and the mathematical computations produced from the inputs. While an expert computer scientist should be able to trouble-shoot the code, they will not be able to interpret the **thousands of numbers used in the weightings and the billions of calculations** from those numbers. This is what is meant when people describe an AI/ML system as a “black box.” Savage N. 2022.

Breaking into the black box of artificial intelligence. Nature, March 22.

<https://www.nature.com/articles/d41586-022-00858-1>

Trusting a “Black Box”

- The opacity of AI/ML systems is a problem because one might argue that we should not use these tools if we cannot trust them, and we cannot trust them if even the best experts do not completely understand how they work. Trust in technology, one might argue, is based on understanding that technology. If we do not understand how a telescope works, then we should not trust in what we see in through the telescope.
- Likewise, if computer experts do not completely understand how an AI/ML system works, then perhaps we should not use the system for important tasks, such as making hiring decisions, diagnosing diseases, analyzing data, or generating scientific hypotheses or theories

Trusting Results

- One way of responding to the “black box” problem is to argue that we do not need to completely understand AI in order to trust it; all that really matters is that it reliably produces correct results.
- Proponents of this view draw an analogy between using AI/ML tools and using other technologies, such as aspirin for pain relief, without fully understanding how they work. All that really matters for trusting a machine, tool, drug is that it works, not that we completely understand how it works.
- **Problem: this is not a very satisfactory response for legal liability, error-analysis (e.g., crashes), product approval (e.g., AI medical devices), and public acceptance.**

Explainable AI

- A second approach, which has been gaining steadily in popularity, is to try to make AI explainable by making it more transparent.
- Disclosure could include:
 - The type, name, and version of AI system used
 - What it was used for
 - How it was used
 - Why it was used
 - Technical details, such as training data, algorithms, models, and optimization procedures, influential features involved in model's decisions, the reliability and accuracy of the system (if known).
- Explainability, according to proponents of this approach, helps to promote trust in AI because it allows users to make rational, informed decisions about using it. Ankarstad A. 2020. What is explainable AI (XAI)? Towards Data Science, April 10. <https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513>

Explainable AI

- The main problem with explainable AI that is may not be explainable to most users because considerable expertise in computer science and/or data analytics may be required to understand the information that is disclosed.
- For transparency to be effective, it must address the audience's informational needs. Explainable AI, at least in its current embodiment, may not address the informational needs of the laypeople, politicians, professionals, regulators (e.g., FDA), judges, jurors, or scientists because the information is too technical. To be explainable to non-experts, the information may need to be expressed in plain, jargon-free language that explains what the AI did and why it did it.

Ethics of Research

- Scientific ethics are norms (i.e., principles, values, or virtues) for the conduct in inquiry.
- These norms apply to many different scientific practices, including research design; experimentation and testing; modelling; concept formation; data collection and storage; data analysis and interpretation; data sharing; publication; peer review; hypothesis and theory acceptance; communication with the public; and mentoring and education.
- Many of these norms are expressed in codes of conduct, professional guidelines, institutional or journal policies, or books and papers on scientific methodology. Others are not formally written down but are implicit in the practice of science.
- Some norms, such as testability, rigor, and reproducibility, are primarily epistemic; while others, such as fair sharing of credit, protection of research subjects, and social responsibility, are primarily moral; while others, such as honesty, openness, and transparency, have epistemic and moral dimensions.

Scientific Norms

- Honesty
- Testability
- Rigor
- Empiricism
- Skepticism
- Explanatory power
- Objectivity
- Realism
- Precision
- Openness
- Transparency
- Reproducibility
- Accountability
- Freedom of inquiry
- Fair sharing of credit
- Confidentiality of peer review
- Collegiality
- Non-discrimination
- Respect for intellectual property
- Protection of human subjects
- Protection of animal subjects
- Safety (physical, biological, psychosocial)
- Stewardship of resources
- Social responsibility

Scientific Norms

- Norms have three sources of justification:
 - To achieve the goals of science
 - To promote collaboration and trust among scientists
 - To promote public trust and accountability
- Norms are more like guidelines than rigid rules; when they conflict, scientists must decide which one should take priority (e.g., openness vs. confidentiality of human data).

Ethical Use of AI in Research

Dealing with Bias

- While reduction and control of bias is widely recognized as essential to good scientific methodology and practice, it takes on added importance in science that uses AI/ML because AI/ML can but also amplify biases inherent in the training data and lend support to policies that are discriminatory, unfair, harmful, or ineffective.
- Moreover, users of AI/ML in research may overconfidently estimate the objectivity of their findings because they are being generated by an “unbiased” machine. The problem of AI bias in medical, psychiatric, and public health research has generated considerable concern, since biases related to race, ethnicity, gender, sexuality, age, nationality, and socioeconomic status in health-related datasets can perpetuate health disparities by supporting biased hypotheses, models, theories, and policies.

Dealing with Bias

- Scientists who use AI/ML in research have special obligations to identify, describe, reduce, control, and correct biases. To fulfill these obligations, scientists must not only attend to matters of research design, data analysis, and data interpretation, but also address issues related to data diversity and representativeness, and interactions between data, algorithms, and applications.
- Scientists must also be accountable for AI bias, both to other scientists and members of the public. To build public trust in AI and promote accountability, and social value, scientists who use AI/ML should engage with affected populations, communities and other stakeholders to obtain their assistance in identifying and reducing potential biases.
- They should explain how and why and AI was used (explainable AI).

Dealing with Error

- Scientists who use AI in their research should disclose and discuss potential sources of AI-related error. Discussion of potential sources of error is important for making research transparent and reproducible.
- Strategies for reducing errors in science include time-honored quality assurance and quality improvement techniques, such as auditing data; validating and testing instruments; and investigating and analyzing random and systemic errors. Replication of results by independent researchers, journal peer review, and post-publication peer review also play a major role in error reduction.
- Accountability requires that scientists take responsibility for errors produced by the use of AIs in research, that they be able to explain why errors have occurred, and that they take necessary steps to correct errors, such as submitting corrections or retractions to the journal.

AI Authorship

- AI authorship became a hot button issue when several papers were published in late 2022 that named LLMs as coauthors.
- Some argued that LLMs could be authors if they make a significant contribution to the research.
- Science magazine stated that not only could AIs not be authors, but they should not be used at all in preparing manuscripts.
- The emerging consensus position seems to be that 1) AIs cannot be authors because they cannot be accountable; 2) they can be used to write or edit papers as long as their use is properly described and disclosed; 3) it is important to give appropriate recognition to the role that an AI has played in research to promote transparency but also so the human authors will not receive more credit than they deserve. Hosseini M, Resnik DB, and Holmes K. 2023. The ethics of disclosing the use of artificial intelligence in tools writing scholarly manuscripts. Research Ethics, June 15.

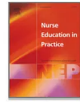
<https://doi.org/10.1177/17470161231180449>

Naming AIs as authors



Nurse Education in Practice

Volume 66, January 2023, 103537



Editorial

Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?

Siobhan O'Connor^a, ChatGPT^b

Show more

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.nepr.2022.103537>

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES



This article was corrected to remove ChatGPT as an author

[Get rights and content](#)

Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models

Tiffany H. Kung, Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng

doi: <https://doi.org/10.1101/2022.12.19.22283643>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

www.oncoscience.us

Oncoscience, Volume 9, 2022

Research Perspective

Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective

ChatGPT Generative Pre-trained Transformer² and Alex Zhavoronkov¹

¹Insilico Medicine, Hong Kong Science and Technology Park, Hong Kong

²OpenAI, San Francisco, CA 94110, USA

Correspondence to: Alex Zhavoronkov, email: alex@insilico.com

Keywords: artificial intelligence; Rapamycin; philosophy; longevity medicine; Pascal's Wager

Received: December 14, 2022 Accepted: December 15, 2022 Published: December 21, 2022

Copyright: © 2022 Zhavoronkov. This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 3.0\)](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Author contributions

ChatGPT produced the majority of the perspective article in response to the query by Alex Zhavoronkov who had a strong desire to publish on the subject. The generated perspective was reviewed by Alex Zhavoronkov who also agreed with the arguments presented by ChatGPT. In response to a direct query regarding co-authorship, **ChatGPT produced multiple arguments why it should not be included as a co-author.** However, due the fact that the majority of the article was produced by the large language model, to set a precedent, the decision was made to include ChatGPT as a co-author and add the appropriate explanation and reference in the article. ChatGPT also assisted with references and appropriate formatting. Alex Zhavoronkov reached out to Sam Altman, the co-founder and CEO of OpenAI to confirm, and received a response with no objections. The ability of the large language models, and other AI systems to make meaningful contributions to the academic work may justify future co-authorship on academic perspective, review and research papers.

Naming AIs as authors?

For

- AIs can make **substantial contributions** to research, including writing and data analysis or interpretation.
- **Credit** should be given where it is due and not give where it is not due.

Against*

- Current AIs cannot clearly explain what they do, how they do it, and why (black box problem) so they cannot be held **accountable**.
- Because current AIs lack consciousness, emotion, self-awareness they are not moral agents and cannot be held **morally responsible**.
- Credit can be given by properly acknowledging AI use.

*Note: these arguments also imply that AIs cannot be listed as inventors on patents applications or hold copyrights.

ICMJE

At submission, the journal should require authors to disclose whether they used artificial intelligence (AI) assisted technologies (such as Large Language Models [LLMs], chatbots, or image creators) in the production of submitted work. Authors who use such technology should describe, in both the cover letter and the submitted work, how they used it. **Chatbots (such as ChatGPT) should not be listed as authors because they cannot be responsible for the accuracy, integrity, and originality of the work, and these responsibilities are required for authorship (see Section II.A.1). Therefore, humans are responsible for any submitted material that included the use of AI-assisted technologies.** Authors should carefully review and edit the result because AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased. Authors should not list AI and AI assisted technologies as an author or co-author, nor cite AI as an author. Authors should be able to assert that there is no plagiarism in their paper, including in text and images produced by the AI. Humans must ensure there is appropriate attribution of all quoted material, including full citations (International Committee of Medical Journal Editors. 2023. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. <https://www.icmje.org/icmje-recommendations.pdf>).

Disclosure for AI writing use

- An evolving topic; more work is needed on disclosure standards.
- Disclose substantial use of AI in research and writing.
 - What AI was used for (e.g., background research, citations, editing, proof reading, data analysis)
 - Which parts of the paper was it used in.
 - What was type of AI, e.g., name and version date.
 - When was it used.
 - What were the prompts used to generate text.

Substantial Use of AI [Resnik and Hosseini, in preparation]

- **The AI tool generates content.** For example, using an AI tool to write sections of a paper, translate language in the paper, or create synthetic data should be disclosed because the AI has generated content, but using an AI tool to edit a paper for grammar or suggest synonyms or phrases need not be disclosed because the tool is not creating content.
- **The AI synthesizes content.** For example, using an AI tool to piece together notes and draft of parts of a paper to create a final version would be a substantial use.
- **The AI tool analyses data or images.** For example, using an AI tool to analyze genomic data, text, or radiologic images would be substantial uses. As discussed above, the rationale for disclosure is similar to the rationale for disclosing other methods and tools used in research, such as statistical software.
- **The AI tool makes a decision that affects the results of the research.** For example, using an AI tool to extract data from articles to do a systematic review would be a substantial use of the tool because the tool would be making data extraction decisions that affect the outcome of the systematic review.



Undisclosed use of AI

According to some estimates, between 1% and 5% of scientific articles published since 2023 include undisclosed AI-generated text.

Andrew Gay. ChatGPT “contamination”: estimating the prevalence of LLMs in the scholarly literature.

<https://arxiv.org/pdf/2403.16887.pdf>

Hu-Zi Cheng, Bin Sheng, Aaron Lee, Varun Chaudhary, Atanas G. Atanasov, Nan Liu, Yue Qiu, Tien Yin Wong, Yih-Chung Tham, Ying-Feng Zheng. Have AI-Generated Texts from LLM Infiltrated the Realm of Scientific Writing? A Large-Scale Analysis of Preprint Platforms. bioRxiv 2024.03.25.586710;

<https://doi.org/10.1101/2024.03.25.586710>



Undisclosed use of AI

“Certainly, here is a possible introduction for your topic: Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities...” The three-dimensional porous mesh structure of Cu-based metal organic-framework - aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries. Manshu Zhang , Liming Wu , Tao Yang , Bing Zhu , Yangai Liu *Surfaces and Interfaces* Volume 46, March 2024, 104081

TORTURED PHRASES FOUND IN COMPUTER-SCIENCE PAPERS 328 | *Nature* | Vol 596 | 19 August 2021

Scientific term phrase

Big data
Artificial intelligence
Remaining energy
Cloud computing
Signal to noise

Tortured

Colossal information
Counterfeit consciousness
Leftover vitality
Haze figuring
Flag to commotion

[100 Papers with Evidence of Undisclosed AI use](https://retractionwatch.com/papers-and-peer-reviews-with-evidence-of-chatgpt-writing/)

<https://retractionwatch.com/papers-and-peer-reviews-with-evidence-of-chatgpt-writing/>

Retraction

- The PLOS ONE Editors (2024) Retraction: A comparative analysis of blended learning and traditional instruction: Effects on academic motivation and learning outcomes. PLoS ONE 19(4): e0302484. <https://doi.org/10.1371/journal.pone.0302484>.
- Concerns were raised about potential undisclosed use of an artificial intelligence tool to generate text in the article due to inclusion of the phrase “regenerate response” and extensive reference list concerns. PLOS was unable to verify 18 of the 76 cited references, and 6 additional references appear to contain errors. The first and corresponding authors stated that the authors were responsible for the manuscript content and that the only AI tool used during manuscript preparation was Grammarly, to improve language. They provided replacement references but several of the replacements did not appear to support the corresponding statements in the article.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

Step1 : The solution of the nonlinear ordinary differential equation (NODE)

$$U(x) = g_0 + \sum_{i=1}^L \left(\frac{Z(x)}{1+Z(x)^2} \right)^{i-1} \left(g_i \frac{Z(x)}{1+Z(x)^2} + f_i \frac{1-Z(x)^2}{1+Z(x)^2} \right), \quad (5)$$

is taken. Here g_i , and f_i are constants ($g_i \neq 0$ or $f_i \neq 0$) to be found later. The following equation exists for the $Z(x)$ function:

$$Z'(x) = \sqrt{s + cZ^2(x) + rZ^4(x)}, \quad (6)$$

also, s , c and r constants are depend m.

Step2 : The value of L is found by the principle of balance.

Step3 : Substituting Eq. (5), with Eq. (6) into Eq. (4), we obtain a polynomial expression that depends on the Jacobi elliptic function $Z(x)$. By equating the coefficients of $Z^l(x)$, $\{l = 0-7\}$ equal to zero, we obtain a system of equations. We solve this system to find the unknown parameters. The solutions of Eq. (5) are represented in Table [1] based on the values of the parameters s , c and r :

Regenerate response

Table 1: Jacobi Elliptic Functions

No.	s	c	r	$Z(x)$
1	1	$-1 - m^2$	m^2	$sn(x)$

Software to Detect AI writing

The screenshot shows the GPTZero website. At the top, there is a navigation bar with the GPTZero logo, links for SOLUTIONS, RESOURCES, PRICING, NEWS, TEAM, LOG IN, SALES, and GET STARTED. The main content area features a large heading: "More than an AI detector Preserve What's Human". Below this, a paragraph states: "Since inventing AI detection, GPTZero incorporates the latest research in detecting ChatGPT, GPT4, Google-Gemini, LLaMa, and new AI models, and investigating their sources." There are five circular profile pictures of team members. A "CREATE FREE ACCOUNT" button with a right-pointing arrow is located below the profiles. On the right side of the page, there is a large, light-colored box containing a text input area. Above the input area, it asks "Was this text written by a human or AI?" and "Try detecting one of our sample texts:". Below this are five buttons: ChatGPT, GPT4, Claude, Human, and AI + Human. The text input area has a placeholder "Paste your text here..." and a character count "0/5,000 characters" with an "UPGRADE" button next to it.

Note: These are examples only. No endorsement is intended.

AI Detector by Grammarly

Navigate responsible AI use with our AI checker, trained to identify AI-generated text. A clear score shows how much of your work appears to be written with AI so you can submit it with peace of mind.

Type or paste your text.



Let's get started

Do LLMs violate copyrights?

Do they plagiarize?

Does using text to train an LLM constitute fair use?

Are the outputs of LLMs copyrightable?

The copyright office has said that you can't

Exhibit J, NY Times vs. OpenAI

Note: Some LLMs allow the user to control the output to minimize the risk of plagiarism.

There is also plagiarism checking software, such as iThenticate and Turnitin

Output from GPT-4:

many of America's other global companies — aren't nearly as avid in creating American jobs as other famous companies were in their heydays.

Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple's contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple's other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronics designers rely upon to build their wares.

Actual text from NYTimes:

many of its high-technology peers — are not nearly as avid in creating American jobs as other famous companies were in their heydays.

Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple's contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple's other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronics designers rely upon to build their wares.

<https://chatgptiseatingtheworld.com/2023/12/27/exhibit-j-to-new-york-times-complaint-provides-one-hundred-examples-of-gpt-4-memorizing-content-from-the-new-york-times/>

Research Misconduct

- Failure to appropriately control AI-related errors could make scientists liable for research misconduct, if they intentionally, knowingly, or recklessly disseminate false data or plagiarize. Although most misconduct regulations and policies distinguish between misconduct and honest error, many do permit misconduct findings based **on recklessness**.
- While the difference between recklessness and negligence can be difficult to determine, one way of thinking of recklessness is that it involves an indifference to or disregard for the veracity or integrity of research. **For example, a person who uses ChatGPT to write a paper and does not carefully to check its work for errors, could be liable for research misconduct.**

Synthetic Data

- Generative AI can create synthetic data for use in modelling, hypothesis development, and piloting and validation of studies.
- It is also possible that some scientists may use AI/ML systems to deliberately fabricate or falsify data or images.
- Although I do not know of any misconduct cases where synthetic data has been passed off as real data, it is only a matter of time until this happens, given the pressures to produce data and the temptations to cut corners.
- Also, using synthetic data in research, even appropriately, may blur the line between **real** and **fake** data and undermine the commitment to honesty and integrity in research (i.e., the slippery slope). So, the situation bears watching. Savage N. 2023.

Synthetic data could be better than real data. Nature. Apr 27

<https://www.nature.com/articles/d41586-023-01445-8>

SCIENTISTS USED CHATGPT TO GENERATE A WHOLE PAPER FROM DATA

An autonomous system prompted ChatGPT to write a paper that was fluent and insightful, but flawed.

By Gemma Conroy

A pair of scientists has produced a research paper in less than an hour with the help of ChatGPT – a tool driven by artificial intelligence (AI) that can understand and generate human-like text. The article was fluent and insightful, but researchers say that there are many hurdles to overcome before the tool can be truly helpful.

The goal was to explore ChatGPT's capabilities as a research 'co-pilot' and discuss its

advantages and pitfalls, says Roy Kishony, a biologist and data scientist at the Technion – Israel Institute of Technology in Haifa.

The researchers designed a software package that automatically fed prompts to ChatGPT and built on its responses to refine the paper over time. This autonomous data-to-paper system led the chatbot through a step-by-step process that mirrors the scientific process, from initial data exploration, through writing data-analysis code and interpreting the results, to writing a polished manuscript.

To put their system to the test, Kishony

The Impact of Fruit and Vegetable Consumption and Physical Activity on Diabetes Risk among Adults

Data to Paper

June 23, 2023

Abstract

Diabetes is a global health concern, and identifying modifiable risk factors is essential for prevention. We investigated the association between fruit and vegetable consumption, physical activity, and the risk of diabetes among adults. Using data from the Behavioral Risk Factor Surveillance System (BRFSS) 2015 survey, logistic regression analysis was conducted, controlling for age, sex, BMI, education, and income. Our results show that higher fruit and vegetable consumption is associated with a reduced risk of diabetes. Moreover, engaging in regular physical activity strengthens this association. This study addresses a gap in the literature by providing evidence on the protective effects of fruit and vegetable consumption and physical activity in relation to diabetes risk. However, limitations, such as self-reported data and potential confounders, should be considered. Our findings highlight the importance of promoting healthy lifestyle behaviors and have implications for diabetes prevention interventions among adults.

Table 1: Association between fruit and vegetable consumption and diabetes risk: Logistic regression results

Variable	Coeff.	Std. Err.	p-value
Intercept	-4.861	±0.050	< 10 ⁻⁴
Fruit & Vegetable	-0.181	±0.012	< 10 ⁻⁴
Age (years)	0.211	±0.002	< 10 ⁻⁴
Sex (Male)	0.329	±0.013	< 10 ⁻⁴
BMI	0.085	±0.001	< 10 ⁻⁴
Education	-0.108	±0.007	< 10 ⁻⁴
Income	-0.147	±0.003	< 10 ⁻⁴

Association between Physical Activity, Fruit and Vegetable Consumption, and Diabetes Risk

To further explore the relationship between fruit and vegetable consumption, physical activity, and diabetes risk, we performed a logistic regression analysis controlling for age, sex, BMI, education, income, and physical activity (Table 2). The results demonstrate that physical activity (Coefficient = -0.211, p-value < 10⁻⁴) and fruit and vegetable consumption (Coefficient = -0.052, p-value = 0.016) are independently associated with a reduced risk of diabetes. Moreover, the interaction term between fruit and vegetable consumption and physical activity is also statistically significant (Coefficient = -0.143, p-value < 10⁻⁴). This indicates that the combined effect of engaging in physical activity and consuming fruits and vegetables is even more protective against diabetes.

The inclusion of physical activity and the interaction term in the logistic regression model improves its predictive power, as indicated by a higher pseudo R-squared value of 0.1263 compared to 0.1242 in the model without the interaction term. These results provide insights into potential mechanisms by which lifestyle interventions, such as increasing fruit and vegetable consumption and engaging in physical activity, may contribute to reducing the burden of diabetes among adults.

The negative correlation coefficient of -0.181 between fruit and vegetable consumption and diabetes risk suggests that for every unit increase in fruit and vegetable consumption, the odds of developing diabetes decrease by

Methods

Data Source

The data for this study was obtained from the CDC's Behavioral Risk Factor Surveillance System (BRFSS), specifically from the year 2015 survey. The BRFSS is an annual health-related telephone survey that collects information on health-related risk behaviors, chronic health conditions, and the use of preventative services from over 400,000 Americans. The dataset used for this study consists of 253,680 responses with 22 features, including diabetes status, fruit and vegetable consumption, physical activity level, and demo-

graphic variables. The dataset was provided as a comma-separated values (CSV) file.

Data Preprocessing

The pre-processing of the data was performed using Python programming language. First, missing values were removed from the original dataset, resulting in a clean dataset of 253,680 responses. This step ensures that the subsequent analysis is conducted on complete data. Next, a new variable called "FruitVeg" was created by combining the "Fruits" and "Veggies" variables using a logical AND operation. This new variable represents whether an individual consumes at least one fruit and one vegetable each day. These pre-processing steps were performed using the pandas library in Python.

Data Analysis

To examine the association between fruit and vegetable consumption, physical activity, and the risk of diabetes among adults, logistic regression analysis was conducted using the statsmodels library in Python. In the first analysis step, a logistic regression model was fitted with the "Diabetes_binary" variable as the dependent variable and "FruitVeg," "Age," "Sex," "BMI," "Education," and "Income" as independent variables. This analysis aimed to determine the association between fruit and vegetable consumption and the risk of diabetes, while controlling for demographic and health-related factors.

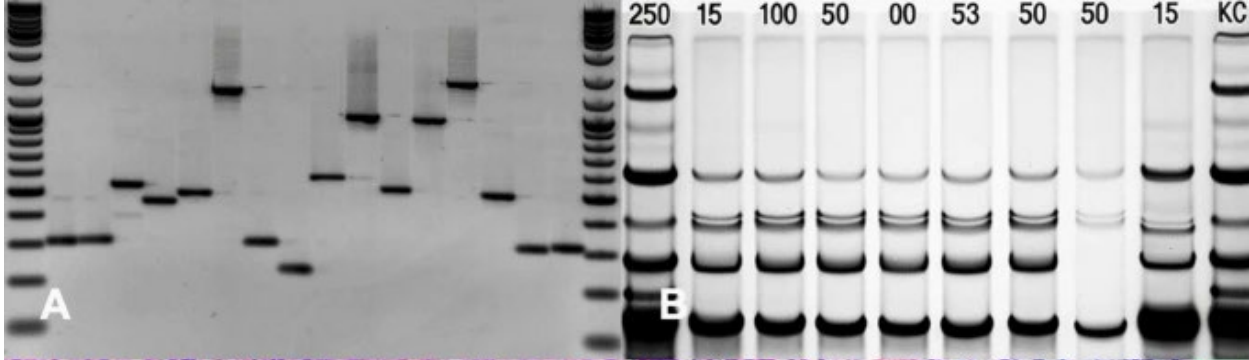
In the second analysis step, an interaction term between fruit and vegetable consumption ("FruitVeg") and physical activity level ("PhysActivity") was introduced in the logistic regression model. The model included the main effects of "FruitVeg" and "PhysActivity," as well as the interaction term "FruitVeg_PhysActivity." This analysis aimed to investigate whether the association between fruit and vegetable consumption and diabetes risk is modified by physical activity level.

The results of the logistic regression analyses, including odds ratios and corresponding p-values, were obtained from the fitted models. Additionally, descriptive statistics for the dataset were calculated using the pandas library. The results were written to a text file named "results.txt" for further examination and reporting.

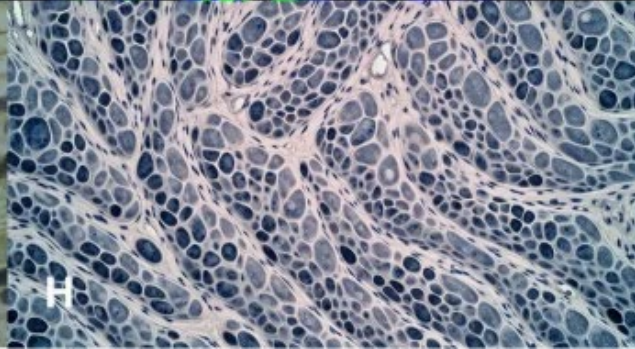
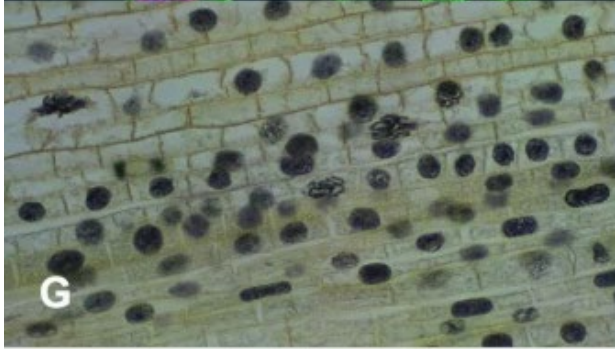
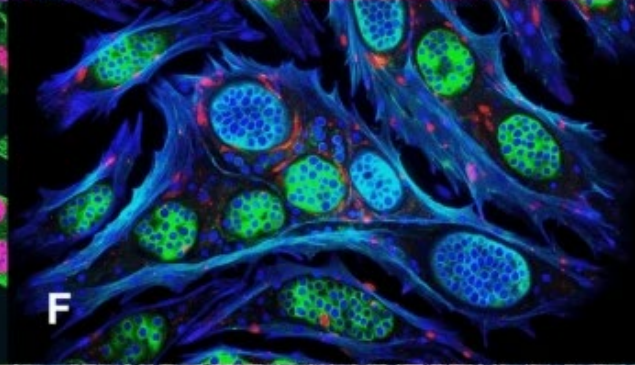
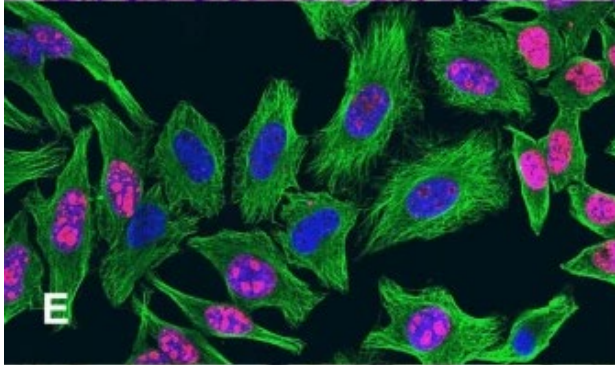
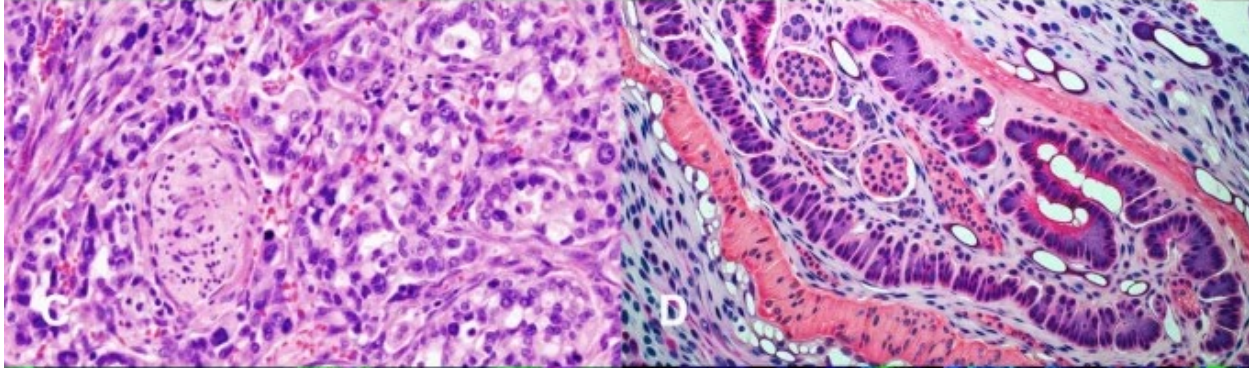
These analysis steps provide insights into the association between fruit and vegetable consumption, physical activity, and the risk of diabetes among adults, while controlling for potential confounding factors.

References

- [1] Pouya Saeedi, Inga Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. Motala, K. Ogurtsova, J. Shaw, D. Bright, and Rhys Williams. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas, 9 th edition. 2019.
- [2] S. Wild, G. Roglić, A. Green, R. Sicree, and H. King. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes care*, 27 5:1047–53, 2004.
- [3] A. Uloko, B. Musa, M. Ramalan, I. Gezawa, F. Puepet, A. Uloko, M. Borodo, and K. Sada. Prevalence and risk factors for diabetes mellitus in nigeria: A systematic review and meta-analysis. *Diabetes Therapy*, 9:1307 – 1316, 2018.
- [4] Xiao-Hua Li, Fei fei Yu, Yu hao Zhou, and Jia He. Association between alcohol consumption and the risk of incident type 2 diabetes: a systematic review and dose-response meta-analysis. *The American journal of clinical nutrition*, 103 3:818–29, 2016.
- [5] A. Herbst, O. Kordonouri, K. Schwab, , F. Schmidt, and R. Holl. Impact of physical activity on cardiovascular risk factors in children with type 1 diabetes. *Diabetes Care*, 30:2098 – 2100, 2007.

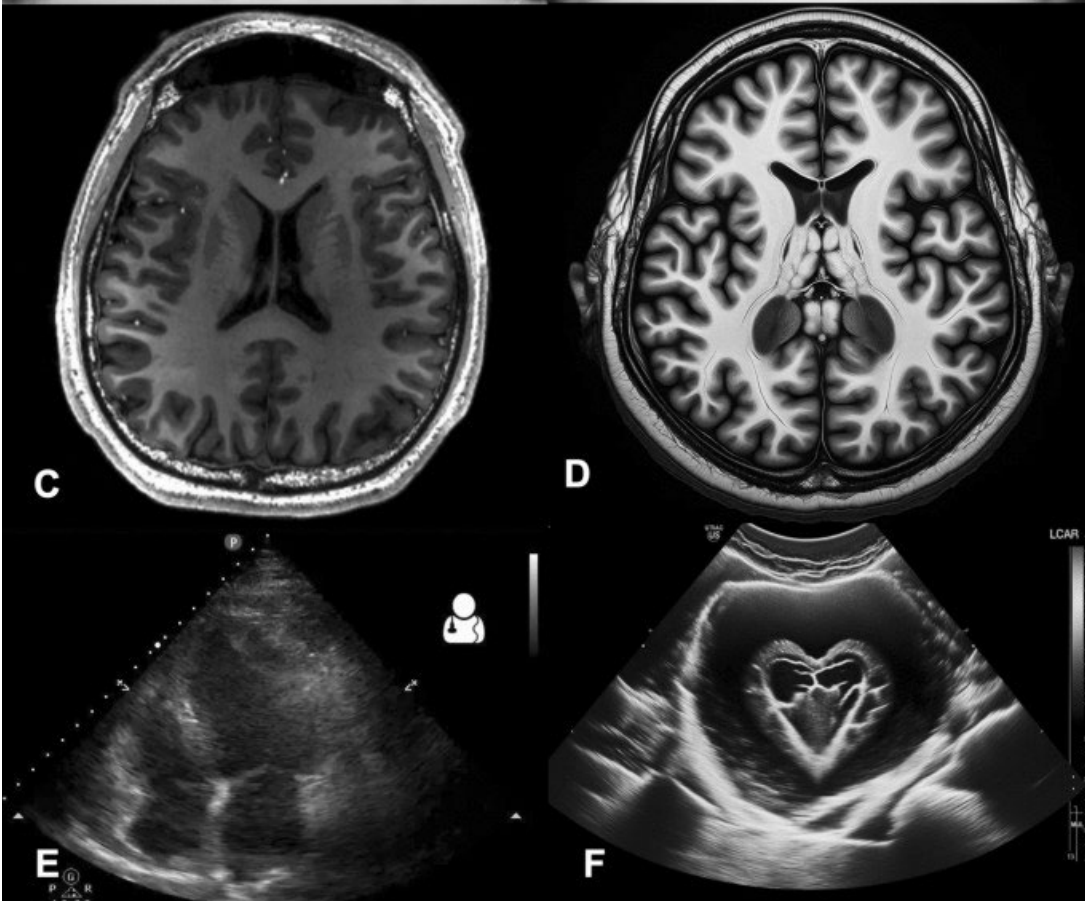
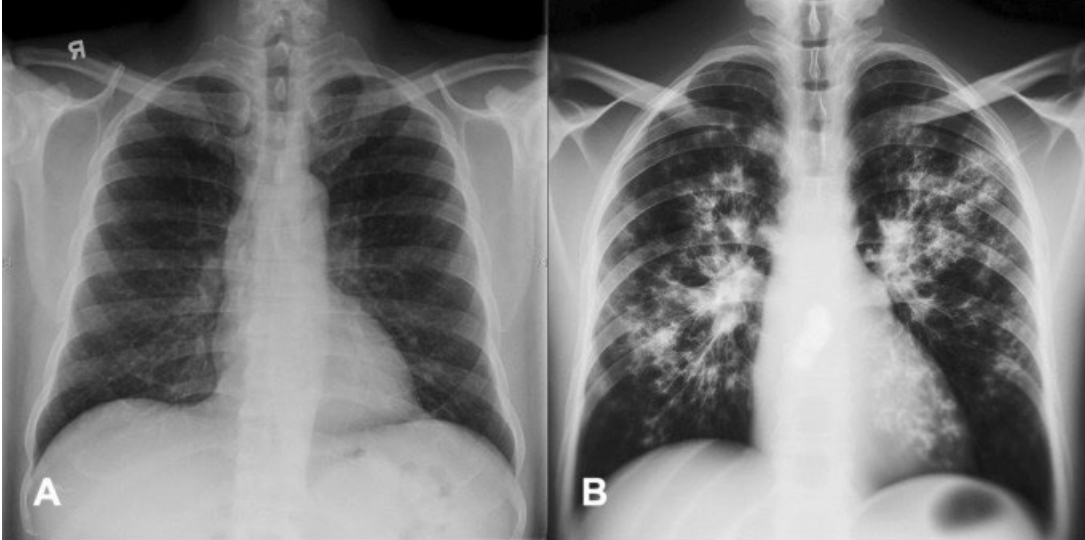


DALL-E-3
western blot and
microscopy
images



Kim, J.J.H., Um, R.S., Lee, J.W.Y. *et al.* Generative AI can fabricate advanced scientific visualizations: ethical implications and strategic mitigation framework. *AI Ethics* (2024).

<https://doi.org/10.1007/s43681-024-00439-0>



DALL-E-3 X-ray and MRI images

Kim, J.J.H., Um, R.S., Lee, J.W.Y. *et al.* Generative AI can fabricate advanced scientific visualizations: ethical implications and strategic mitigation framework. *AI Ethics* (2024).
<https://doi.org/10.1007/s43681-024-00439-0>

AI-enabled image fraud in scientific publications





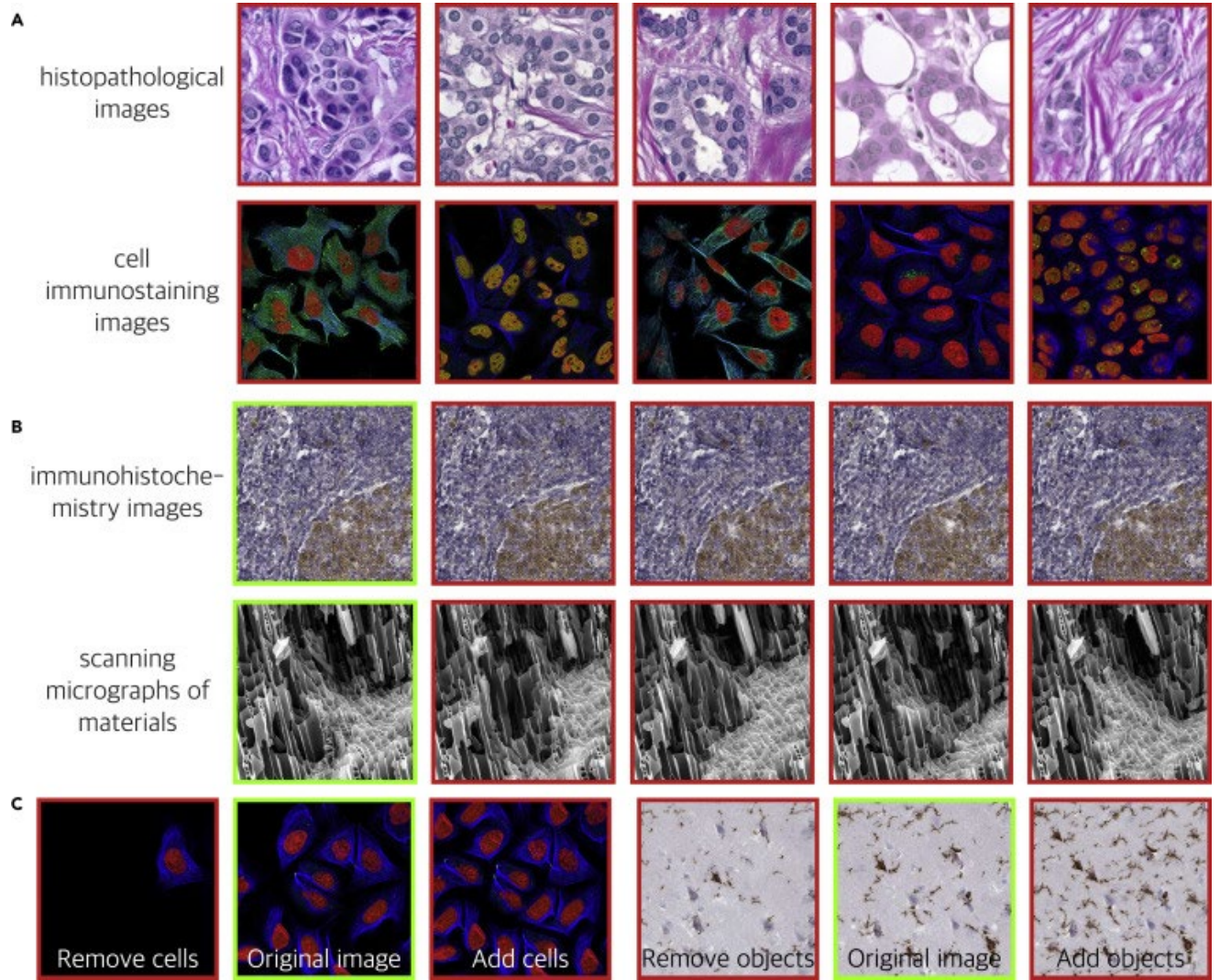
Jinjin Gu¹, Xinlei Wang¹, Chenang Li², Junhua Zhao^{2,3}  , Weijin Fu⁴  , Gaoqi Liang², Jing Qiu¹

Figure 1. Scientific image fraud by intelligent models
 We show several fake images generated by generative models. The images with the red border are all computer generated, while the images with the green border are real ones.



Dealing with Fake Data and Images

- Use AI tools to detect fake data/images
- Watermark synthetic data
- Certify real data (for example, a certification stamp linked to the data, protected by block chain and other security methods—this would be very expensive, cumbersome, and limited).
- **Technical solutions can go but so far, and we must rely on human solutions—education, trust.**

Confidentiality

- The use of AI/ML in research, especially the use of LLMs, such as ChatGPT, raises issues related to the privacy and confidentiality of data.
- ChatGPT, for example, stores the data submitted by users, including data submitted in initial prompts and subsequent interactions with the LLM. The data may also be used to train the AI.
- It is possible that other users of the system could gain access to the data.
- Due to concerns about breaches of confidentiality, the National Institutes of Health (NIH) recently decided to prohibit the use of generative AI technologies, such as LLMs, in grant peer review. Researchers who use an LLM to edit a document should assume not assume the confidentiality is protected, **unless the LLM is a local instance of AI behind by an institutional firewall and other security measures.**

Big picture concerns



De-skilling of human beings, losing writing skills, less emphasis on scientific writing if a machine can do it.



Because writing and thinking are connected, loss writing skills can lead to degrading of scientific thinking skills; turning over thinking to a machine.



Loss of scientific jobs to AI.



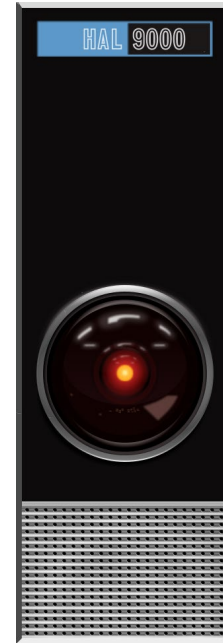
Loss of creativity and diversity in scientific writing.



Environmental impact.

Future AIs

- In the future, scientists and engineers may develop AIs that can intelligibly explain their own behavior. However, even if an AI can explain its own behavior, we still may not consider it to be morally responsible for its behavior.
- Moral agency requires the capacity to perform intentional (or purposeful) actions, the capacity to understand moral norms, and the capacity to make decisions based on moral norms. These capacities also presuppose additional capacities, such as consciousness, self-awareness, personal memory, perception, general intelligence, and emotion.
- While computer scientists are making some progress on developing AIs that can make decisions based on moral norms, they are still a long way from developing AIs with genuine moral agency.



https://upload.wikimedia.org/wikipedia/commons/thumb/2/2e/Hal_9000_Panel.svg/330px-Hal_9000_Panel.svg.png



https://en.wikipedia.org/wiki/Data_%28Star_Trek%29

Conclusion: Recommendations for Ethical Use of AI in Research

Recommendation	Normative Justification
Researchers and software developers are responsible for identifying, describing, reducing, and controlling AI-related biases and random errors.	Accountability, objectivity, reproducibility, rigor, transparency, honesty, social responsibility, fairness
Researchers should disclose, describe, and explain their use of AI in research in language that can be understood by non-experts.	Accountability, transparency, reproducibility, rigor, objectivity, social responsibility, fairness
If appropriate, researchers should engage with relevant communities, populations, or stakeholders concerning the use of AI in research to obtain their advice and assistance and address their interests and concerns.	Accountability, transparency, social responsibility, rigor, fairness
Researchers may be liable for misconduct if they intentionally, knowingly, or recklessly use AI to fabricate or falsify data or commit plagiarism. AI synthetic data use should be appropriately explained and labelled.	Accountability, honesty, reproducibility, rigor

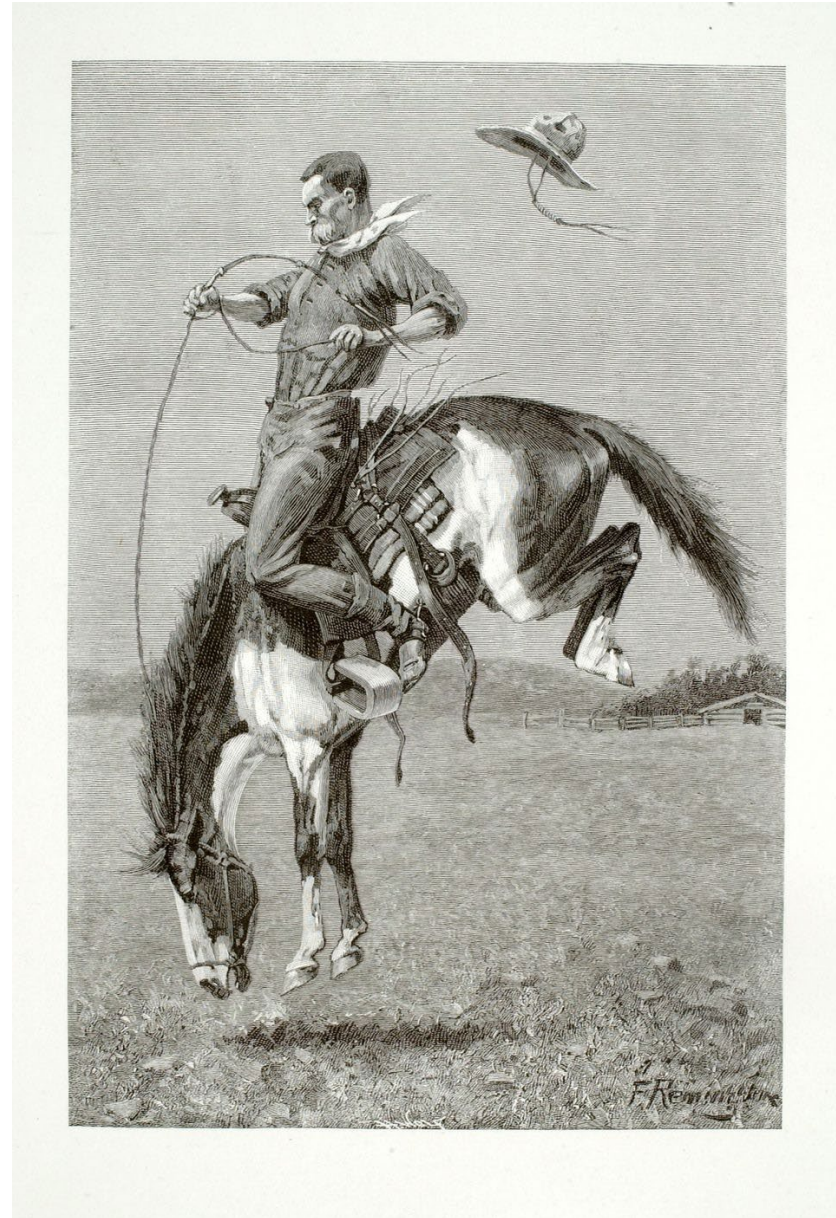
Conclusion: Recommendations for Ethical Use of AI in Research

Recommendation	Normative Justification
<p>AI systems should not be named as authors, inventors, or copyright holders but their contributions to research should be disclosed and described.</p>	<p>Honesty, transparency, accountability, fair attribution of credit, collegiality</p>
<p>AI systems should not be used in situations that may involve unauthorized disclosure of confidential information related to human research subjects, unpublished research, potential intellectual property claims, or proprietary or classified research.</p>	<p>Protection of human subjects, protection of intellectual property, confidentiality of peer review, social responsibility</p>
<p>Education and mentoring in responsible conduct of research should include discussion of ethical use of AI.</p>	<p>Accountability, reproducibility, rigor, social responsibility, honesty, transparency, fair attribution of credit</p>

Final Thoughts

- AI is a highly disruptive technology that presents opportunities and dangers.
- AI use will be like the wild west for a while until policies and best practices emerge and AI use becomes normalized.
- Hang on—you're in for a wild ride!

<https://americanart.si.edu/artwork/bucking-bronco-27919> A Bucking Bronco, Henry Wolf after Frederick Remington, Smithsonian Institution



Acknowledgments

- Mohammad Hosseini, PhD, Northwestern University

- Presentation based on:

Resnik DB and Hosseini M. The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI Ethics* (2024). <https://doi.org/10.1007/s43681-024-00493-8>. Published online May 27, 2024.