Publication status: This preprint has been published elsewhere.
DOI of the published preprint: https://doi.org/10.1590/1808-16570000422026

Don’t Give It Away: The Hidden Risks of Generative AI for Scientists

Anarosa Brandão

https://doi.org/10.1590/SciELOPreprints.15357

Submitted on: 2026-03-09
Posted on: 2026-03-13 (version 1)
(YYYY-MM-DD)

ASSAY

Don’t Give It Away: The Hidden Risks of Generative AI for Scientists

Brandao, Anarosa

https://orcid.org/0000-0001-8992-4768

anarosa.brandao@usp.br
Escola Politécnica – Universidade de São Paulo

Abstract
This text aims to provide readers with an overview of the artificial intelligence that powers conversational agents and has become commonplace in recent years, after the launch of ChatGPT[1]. Also, we outline some benefits and potential risks associated with the misuse of this technology in the scientific research ecosystem. Finally, we suggest ways to mitigate these risks.
Keywords: Artificial Intelligence, ChatGPT, scientific integrity

Artificial Intelligence (AI) is a field of Computer Science that emerged in the 1950s. The term was coined in 1955 in the proposal drafted by McCarthy, Minsky, Rochester, and Shannon for the *Dartmouth Summer Research Project on Artificial Intelligence* [1] to organize a workshop during the summer of 1956 at Dartmouth College in Hanover, USA. The aim of the workshop was to discuss questions related to what McCarthy and colleagues called artificial intelligence issues. That summer, discussions covered, among other topics, how computers could use natural language from a word-manipulation perspective and what the formal modeling of a neural network would look like. Since then, AI has gone through phases of euphoria and despair, mostly due to the processing limitations of computers.

Currently, in addition to the exponential increase in processing power, new AI techniques for text analysis have made it possible for computers to use natural language effectively. This was one of the questions raised in the summer of 1956 and is properly answered by now. This advancement in AI has enabled the penetration of intelligent services throughout society, increasing people's perception of how much this technology can benefit them. This perception, combined with the integration of intelligence into several services provided to society — such as mobile communication services and social networks — causes the public to quickly adopt technology without understanding that there are also risks associated with its use.

Naturally, the ecosystem of scientific research can benefit from its use. Nevertheless, additional care is required to prevent AI misuse from affecting its processes, leading to unpredictable and damaging results. To better understand the risks, it is important to identify which intelligence we are referring to...

[1] https://chatgpt.com

Given the very characteristics of AI, there is no single definition for the term. Russell and Norvig [2] adopt the idea of an "agent" to define the intelligence attributed to computational systems. In this case, an agent is an entity that possesses sensors and actuators and is immersed in an environment. Its intelligence lies in how this entity processes the information collected from the environment by its sensors and in the actions resulting from this processing, which can modify the environment through the actuators. Thus, the authors organize AI definitions into four dimensions:

- intelligence underlying systems that act like humans;
- intelligence underlying systems that reason like humans;
- intelligence underlying systems that reason rationally;
- intelligence underlying systems that act rationally.

When we talk about an intelligence that reasons and acts rationally, we refer to intelligences that allow the agent to use logic or rules to think and justify its actions in deterministic situations or, in the presence of uncertainty, through probabilistic models. In this case, to analyze the reasons that led the agent to take certain decisions are explainable.

On the other hand, when the intelligence in question is one that acts like a human being, we refer to an intelligence capable of communicating in natural language, storing acquired knowledge, reasoning, learning, and inferring new things using such knowledge, perceiving the world, and manipulating its objects. To think like humans, this intelligence needs to acquire cognitive capacity that mimics human ability. In this sense, several areas of study have been dedicated to researching how to perform each of these tasks like humans, namely: natural language processing, knowledge representation, automated reasoning, machine learning, computer vision, and robotics.

Despite this separation into areas of study, machine learning is a field that permeates the others, particularly computer vision, natural language processing, and automated reasoning. In fact, since the early 2000s, we have witnessed an unprecedented evolution of machine learning techniques, driven by increased processing capacity and the vast availability of data to feed the learning process. These techniques were primarily dedicated to solving tasks related to image and text processing. Essentially, underlying these techniques there are pattern recognition and the probabilistic inference of the next pattern, without any semantic understanding of the processed content. Furthermore, patterns are learned from data collected in the past — which may be recent, but is still the past.

Beginning in mid-2017, with the publication of the article "Attention is All You Need"[2] by a group of Google researchers \cite{vaswani2023attentionneed}, a new way of processing text was presented, making text translation tasks faster and more accurate. From then on, intelligent solutions based on this form of processing were created for other tasks associated with natural language text processing. For this evolution to occur, massive amounts of text had to be used to train these solutions, leading to the creation of {\bf Large Language Models (LLMs)}. These models are the foundation of intelligent services currently available

[2] referring here to the 2023 revision

to the general public, such as ChatGPT, Claude[3], Gemini[4], and Perplexity[5]. An important note is regarding the energy and environmental costs associated with these services. Training these LLMs and maintaining the services associated with their use require significant processing, leading to a high consumption of energy and water (needed to cool the machines). But that could be the subject of another essay...

Returning to intelligent services, they offer users an interface that receives general questions in natural language and returns answers in natural language that are grammatically correct, contextually consistent, **but not necessarily accurate** text. This response is generated based on the patterns learned by the model and the context in which it is placed. The underlying AI of this type of service generates responses that can repeat existing patterns, present them in combined ways, or infer new text content, which in turn can be transformed into various formats, such as image, audio, or video. This AI is called *Generative AI*, a type of machine learning that uses a technique called deep neural networks to simulate the learning and decision-making processes of the human brain. This is why it is possible to dialogue with agents like ChatGPT and feel as though the answers could have been crafted by human beings. And that is where the risks arise, and why I continue to advocate for the need to train people for the (proper) use of AI...

Since knowledge evolves based on existing knowledge, LLMs have evolved from the knowledge generated through human interactions with systems like ChatGPT by prompting questions or requests. The business model of big techs consists of providing their services *for free* in exchange for using all the content of the prompts and its associated results. This means that while prompting with ChatGPT in a free account, all the knowledge from the conversation will be used to improve the LLM that feeds the knowledge of ChatGPT.

Considering the scientific research ecosystem, the adoption of AI services to support its methods and processes is a reality that must be continuously monitored considering scientific integrity. In this ecosystem, a fundamental part is the writing and dissemination of research results, since new knowledge is highly dependent on existing knowledge. Understanding how Generative AI works and the big techs business model reveals that some phases of the scientific writing and publishing process may provide content to improve these LLMs even before they are published in appropriate venues. This occurs whenever a free service is used, for example, to improve a text to fit a journal's style or even to translate it into another language. Using these services at an institutional account can prevent this, as such service agreements usually include clauses of DPA - data process agreement. Such clauses ensure privacy and property of all customer data, including prompts and associated answers, in addition to compliance with data regulation. Nevertheless, AI can improve research results by supporting the literature review, text summarization, and hypothesis refinement, among others.

This text is an alarm to avoid using AI services to support scientific research without due diligence. In fact, all free of charge service has a price, and it is exactly all data you prompt

[3] https://claude.ai/
[4] https://gemini.google.com
[5] https://www.perplexity.ai/

on it. Be aware: do not give for free something that you spent a lifetime to create before having the credits for it. Don´t give it away!

References

[1] https://chatgpt.com
[2] referring here to the 2023 revision
[3] https://claude.ai/
[4] https://gemini.google.com
[5] https://www.perplexity.ai/

AUTHORSHIP CONTRIBUTION (CONTRIBUIÇÃO DE AUTORIA)

Brandão, Anarosa

Conceptualization; Resources; Validation; Visualization Writing – Original Draft; Preparation Writing – Review & Editing

AVAILABILITY OF DATA AND MATERIAL (declaração de disponibilidade de dados de pesquisa)

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING

Not applicable.

CONFLICTS OF INTEREST

All authors declare that they have no conflict of interest.

ETHICAL APPROVAL

Not applicable.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a Creative Commons CC-BY license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.