Original Article

Philosophy of medicine meets AI hallucination and AI drift: moving toward a more gentle medicine

"Facilis descensus averno: noctes atque dies patet atri ianua ditis; sed revocare gradium superasque evadere ad auras, hoc opus his labor est."[1] Aeneid, 128-129

Hamidreza Namazi[1], Mohammad Mahdi Radfar[2*]

1. Assistant Professor, Medical Ethics and History of Medicine Research Center, Tehran University of Medical Sciences, Tehran, Iran.
2. Researcher, Medical Doctor, Medical Ethics and History of Medicine Research Center, Tehran University of Medical Sciences, Tehran, Iran.

Abstract

The contemporary world is profoundly shaped by technological progress. Among the advancements of our era is the proliferation of artificial intelligence (AI). AI has permeated every facet of human knowledge, including medicine. One domain of AI development is the application of large language models (LLMs) in health-care settings. While these applications hold immense promise, they are not without challenges. Two notable phenomena, AI hallucination and AI drift, pose setbacks. AI hallucination refers to the generation of erroneous information by AI systems, while AI drift is the production of multiple responses to a single query. The emergence of these challenges underscores the crucial role of the philosophy of medicine. By reminding practitioners of the inherent uncertainty that underpins medical interventions, the philosophy of medicine fosters a more receptive stance toward these technological advancements. Furthermore, by acknowledging the inherent fallibility of these technologies, the philosophy of medicine reinforces the importance of gentle medicine and humility in clinical practice. Physicians must not shy away from embracing AI tools due to their imperfections. Acknowledgment of uncertainty fosters a more accepting attitude toward AI tools among physicians, and by constantly highlighting the imperfections, the philosophy of medicine cultivates a deeper sense of humility among practitioners. It is imperative that experts in the philosophy of medicine engage in thoughtful deliberation to ensure that these powerful technologies are harnessed responsibly and ethically, preventing the reins of medical decision-making from falling into the hands of those without the requisite expertise and ethical grounding.

***Corresponding Author**

Mohammad Mahdi Radfar

Address: *No. 8, Bahare Dead-end, East Brazil Street, Vanak Sq., Tehran, Iran.*
Postal Code : *1435713836*
Tel: *(+98) 21 66 41 96 61*
Email: *mhrad24@gmail.com*

Received: *14 Sep 2024*
Accepted: *25 May 2025*
Published: *27 May 2025*

Citation to this article:
Namazi H, Radfar MM. Philosophy of medicine meets AI hallucination and AI drift: moving toward a more gentle medicine. J Med Ethics Hist Med. 2025; 18: 2. https://doi.org/10.18502/jmehm.v18i2.18812

Keywords: *AI hallucination; AI drift; Data schizophrenia; Medical philosophy; Large language models; Gentle medicine.*

Copyright © 2025 Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International license https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

[1] The gates of hell are open night and day. Smooth the decent and easy is the way, but to return and view the cheerful skies, in this task and mighty labor lies. The quote is a statement from Aeneid by Virgil.

Introduction

Our world today is deeply influenced by novel technological advancements. The advent of artificial intelligence (AI) is a paradigm shift comparable to the invention of the microprocessor, the personal computer, the Internet, and the smartphone. AI has the potential to revolutionize the way we work, learn, travel, receive healthcare, and communicate with each other. Entire industries will be reorganized around AI, and businesses will be differentiated by their ability to harness its power. Whether or not we are comfortable with this paradigm shift, our future will be inextricably linked to AI. No field is immune to its influence, including medicine. A recent innovation in the field of artificial intelligence (AI) is the development of large language models (LLMs) known as Chatbots. Chatbots are electronic systems that simulate human conversations using natural language processing and can be integrated into various platforms like websites and apps. Advanced models like ChatGPT, Gemini and Claude2 are designed to assist users across various fields, from customer service to scientific research, offering capabilities such as answering questions and analyzing data. AI hallucination and AI drift are two key challenges of large language models (LLMs). Hallucination occurs when an AI generates incorrect or fictitious responses, while drift refers to inconsistencies in answers to the same question when posed by different users or as a result of variations in how the question is asked. These are caused by a variety of issues, including the inability of large language models to distinguish between correlation and causation, which will be explained in detail in this article. This is where the inherent role of uncertainty in medical interventions comes into play and factors like evidence quality, clinician experience and patient circumstances shape confidence in diagnoses and treatments. Ultimately, uncertainty is intrinsic to medical practice, highlighting the need for humility and a nuanced approach in addressing each patient's unique situation. The concept of gentle medicine encourages a humble approach in medical practice, advocating for caution in diagnoses and treatments while emphasizing disease prevention. It promotes effective interventions but supports a more restrained application. There is a recognition of the

limitations of artificial intelligence in medicine, highlighting the inherent uncertainties in decision-making. The philosophy of medicine underscores the need for humility when using AI, acknowledging its dual nature as both helpful and potentially harmful. Therefore, medical institutions should prioritize ethical training for professionals to responsibly navigate the integration of AI tools in healthcare.

What is the role of chatbots in medicine?

A chatbot is an electronic system (generally a software) that simulates conversations by responding to keywords or phrases it recognizes and can be integrated into various platforms, such as websites, mobile apps, and messaging platforms. A chatbot is a computer program that simulates human conversation through natural language processing (NLP). Chatbots are often used in customer service applications, but they can also be used for education, entertainment, and public purposes (1).

NLP is a field of computer science that deals with the interaction between computers and human (natural) languages. NLP algorithms are used to understand and generate human language, which is essential for chatbots to simulate human conversation (15).

Chatbots can be integrated into a variety of platforms, including websites, mobile apps, and messaging platforms. This allows users to interact with chatbots in a convenient and natural way.

In this article, we provide a concise overview of several pioneers in the field of artificial intelligence (AI):

1) ChatGPT

ChatGPT, a sophisticated tool developed by OpenAI, has emerged as a prominent player among the diverse array of natural language processing models. This AI-powered chatbot is meticulously crafted to simulate human-like conversations. ChatGPT's functionality stems from its utilization of advanced algorithms programmed to grasp natural language inputs and generate corresponding responses. These responses can be either pre-written or freshly generated by the AI itself (14). Also, OpenAI recently introduced a new search engine empowered by its large language model known as SearchGPT, which can produce answers based on online data from different sources. ChatGPT's utility has been demonstrated across a spectrum of fields, including medical research and literature review. It can effectively assist medical researchers and scientists in crafting articles and abstracts, compressing data or information, providing structural, referential and

titular suggestions, and enhancing the readability of texts. Remarkably, ChatGPT can even generate a complete draft of a paper (10).

While incapable of generating novel ideas, artificial intelligence can effectively organize and refine those conceived by researchers, producing an initial draft. However, this should be considered a starting point for human-led text development, as the AI-generated text falls short of being a substitute for human expertise (8).

2) Gemini

Another promising large language model is Gemini, the Google's AI project that has been launched based on the Palm2 language model. Gemini can search through websites and is empowered by Google search engine, as stated by the Google CEO (3).

You can ask Gemini to answer math problems, schedule a meeting in your workplace, and write an email in a foreign language, choosing your favorite tone, wording and so on. Gemini can be used when online medical information is needed, for instance it can help students learn about up-to-date medical approaches. Furthermore, you can ask Gemini to summarize articles in any way you need (19).

3) Claude2

Claude2 is another large language model introduced by Anthropic, presently available only in the UK and USA. Numerous modifications have been made to Claude2 in order to make it better and faster than the previous large language models. For example, you can upload information to it in the form of a text and ask for analysis. This could be used in medical judgment in complicated cases where analyzing a vast amount of information is hard for the human mind (12).

4) Replika

There are some other types of AI large language models that are not merely for academic uses. Replika is an AI-based language model that you can use on your cellphone as a friend. First, it asks you some personal questions to learn about your personality and afterward, it will treat you like a friend. Just like Samantha in the movie named "Her".

Despite its limitations, Replika is a valuable tool for people who are looking for companionship, support, or simply someone to talk to. It can also be a helpful tool for people exploring their identity and looking to learn new things about themselves (13). In addition to its common application, Replika can be utilized in medicine, for instance medical students can use this tool to strengthen their empathy and communication skills with patients.

Large language models are advancing daily, with new models being designed and implemented to meet emerging needs. The models released by Meta can be cited as examples. These language models are continuously expanding, but they will not be discussed here in order to avoid lengthening this article.

LLMs have had significant success in a variety of tasks, including medical research. Some experts believe that the invention of LLMs and their application in medicine has the potential to revolutionize the field, just as the invention of radiology imaging did (20).

However, it should be noted that LLMs are not without limitations and like radiology imaging, they can be prone to bias and error. Therefore, it is important to carefully evaluate the output of LLMs before relying on them for medical decision-making.

AI hallucination and AI drift: The Siamese Twins of Large Language Models

Two common challenges associated with large language models (LLMs) are AI hallucination and AI drift. Authors tend to call these challenges "schizophrenic reactions". This metaphor helps us to better understand how humane these challenges are.

A large language model hallucination is a situation when an AI model generates senseless or imaginative content. For example, when AI is asked about a scientific concept, the answer may be incorrect and sometimes even completely fictitious. In fact, quite often it will produce false data and responses (4). This phenomenon is referred to as "AI hallucination".

Another term that needs to be discussed is "AI drift", which is a common enough problem with large language models. AI drift happens when one single question leads to different answers. This different answer may arise either when the question has been asked by a different user, or by the same person in another manner or at a different time (6).

Why do these phenomena happen?

Researchers have found some of the main reasons for these alterations of reality. Here are some of the most probable causes:

1) Being outdated

There is a lag on data insertion in many of these large language models. The previous version of ChatGPT, for example, strives to generate output that aligns with anticipated outcomes based on its training data. Due to ChatGPT's knowledge cutoff date of September 2021, it is unable to engage in discussions regarding research papers published in 2024. As a result, any forward-looking statement

extending beyond September 2021 is likely to be inaccurate, constituting a fabrication devoid of any connection to the 2024 publication. ChatGPT's output solely consists of sentences constructed upon the relationship among preceding words, sometimes resulting in nonsensical content and potentially causing a range of consequences, from misunderstandings to the dissemination of misinformation (2).

2) Disconnection from open-source information

There have been reports of ChatGPT not being able to search the web and access the National Library of Medicine or other relevant sources with up-to date information (2). However, this issue has been partially resolved in the newer versions of large language models. As mentioned above, Gemini's advantage over ChatGPT was its access to online databases and real-time articles, but SearchGPT addressed this issue by being able to search online databases.

3) Invalidation of Sources

In ChatGPT and other large language models (LLMs), artificial hallucinations may occur due to a lack of source attribution. The source of specific responses is often unknown, and using different sources with varying information can lead to different answers and conclusions (2).

4) The Role of temperature

The temperature of a large language model (LLM) significantly influences its output and the likelihood of artificial hallucinations. Temperature can be conceptualized as the LLM's level of certainty in its most probable response. An elevated temperature generates a less confident and more diverse response, which can lead to hallucinations. In the context of LLMs, temperature serves as a hyperparameter that controls the degree of stochasticity in the model's output. This parameter ranges from 0 to 1, with a higher temperature producing more diverse and random outputs, while a lower temperature yields more predictable and conservative outputs. Specifically, when the temperature is high, the LLM is less certain about the most likely next word and more prone to selecting less probable words. This can lead to more creative and unexpected responses, but also increases the risk of producing nonsensical or hallucinatory text. Conversely, when the temperature is low, the LLM is more confident in its most likely next word and less likely to deviate from the most probable sequence. This results in more predictable and conservative outputs, but also limits the LLM's ability to generate creative and original texts (4).

Consequently, the temperature parameter serves as a pivotal element in harmonizing the creative and

coherent aspects of an LLM's output. By manipulating the temperature, users can deftly adapt the model's behavior to suit specific tasks, encompassing the generation of creative text formats and the production of factual and reliable information.

The optimal temperature for an LLM is context-dependent and varies in accordance with the specific task at hand. For instance, if an LLM is employed to generate creative text formats such as poems or codes, a higher temperature setting may be preferred to encourage the model to exhibit greater innovation. Conversely, if an LLM is utilized to generate factual text such as a news article or a scientific paper, a lower temperature setting may be employed to guarantee the accuracy of the generated output. For example, ChatGPT uses a temperature of 0.7 for its predictions, which allows it to generate more diverse responses, but also makes it more likely to hallucinate (4).

5) Keywords and prompts

There is an important difference between a research engine motor such as Google and a large language model such as ChatGPT. A search engine uses keywords provided by users to investigate present articles containing them. On the contrary, an AI language model uses prompts given by users to trawl information from the vast number of ideas living within it. When it comes to prompts and ideas, user's prompt semantically precedes the AI. A prompt provided by the user appears to give meaning to distorted, complex and entropic ideas living within that large language model.

This could be another reason for AI drifting. When you ask Dr. House (the main character of the series *House, M.D.*) for the best treatment for your rash, he explores the vast amount of his knowledge and prophetically answers, "Get a divorce!" These AI language models angle the idea by the prompt that you give to them. Hence, different answers can arise when the same question has been asked.

Returning to our metaphor of a schizophrenic person, Dr. House may give a seemingly random answer that would sound nonsensical to an observer who does not understand his thought process. However, this does not necessarily mean that random answers should be treated as low-quality answers.

In the context of large language models (LLMs), randomness can be used to generate more creative and interesting outputs. For example, an LLM can be configured to generate texts with a higher temperature, which will result in more random and diverse outputs. This can be useful for tasks such as generating poems, codes, or other creative content.

6) Correlation and causation dichotomy

There is an obvious distinction between correlation and causation. For instance, I have noticed that over the past month, whenever I ordered my coffee in a red shirt, the weather was cloudy. However, this does not mean that the weather is tuned to my fashion choices! On the other hand, whenever I oversleep, I get to work late. In this case, I can assume that there is a causation between delayed awakening and being late for work.

Correlation and causation, while often intertwined in scientific discourse, represent distinct conceptual frameworks. Correlation quantifies the statistical association between two variables, while causation establishes a causal relationship, implying that one variable directly influences the other.

The philosophical exploration of causation dates back to ancient Greece. In his seminal work, Metaphysics, Aristotle proposed four fundamental causes responsible for natural phenomena: material, formal, efficient, and final. This framework represents the culmination of pre-Socratic and Platonic discussions on causation. The material cause refers to the substance or matter from which an object is composed, while the formal cause embodies the blueprint or design that determines its form and structure. The efficient cause identifies the agent or process that brings the object into existence, and the final cause delineates the purpose or function for which it exists (11).

In the realm of natural sciences, understanding the nature of causation is crucial for elucidating causal relationships and connections among natural phenomena. Scientists must be able to distinguish between entities that genuinely cause natural phenomena and those that merely exhibit a statistical association. Distinguishing correlation from causation is essential for avoiding spurious relationships and drawing valid conclusions about natural phenomena. While correlation may suggest a connection between two variables, it does not establish a causal mechanism. Correlation can arise from a common underlying cause, chance, or even a third, unidentified variable. Establishing causation requires a more rigorous approach, often involving controlled experiments, observational studies, and careful consideration of alternative explanations. By isolating the variable of interest and manipulating it under controlled conditions, scientists can determine whether a causal relationship exists (11).

The example of ordering coffee in a red shirt on cloudy days is an instance of a correlation. There is a statistical relationship between the two variables, but it is not clear whether one causes the other.

There could be a third variable, such as my mood, which is causing both of the other variables.

The example of oversleeping and being late for work, however, represents causation, since it is more likely that the former is causing the latter, rather than vice versa.

It appears that AI language models are better in realizing the correlations than causations. When training an AI language model, machine learning is based on a huge amount of data given to the computer. What happens if the data are merely correlated and not caused by each other? Does this make answers of the AI language model vague? How much would the certainty of the answers be under the circumstances?

The widespread use of data and the rise of big data have given us new capabilities that could usher in a new era. However, they can also lead to the misuse of information, known as "garbage in, garbage out", or GIGO (18). When users feed a large language model with raw, random or rubbish information, the output cannot be expected to be any better.

Large language models are not yet able to distinguish between correlation and causation. Therefore, developers should be cautious when using online data such as tweets and blogs. Security concerns may also increase, given that millions of dollars are now being spent by organizations and governments to produce false information.

Now, is there a moral responsibility for developers and engineers to protect large language models against this type of information? That certainly seems like an important question, since so much is at stake in this regard.

Need for certainty and tolerance of ambiguity

Here we are speaking of "data schizophrenia". When AI has its own brain and neural network, is it not right to entitle AI hallucination as a schizophrenic reaction? Both AI hallucinations and schizophrenic hallucinations can be caused by problems with the way that the brain processes information. Additionally, both can appear to be real or unreal.

In light of the challenges and interpretations discussed above, a new question arises: What is the role of certainty in medical decision-making?

The level of certainty with which a clinician establishes a diagnosis and formulates a treatment plan for a patient is a pivotal aspect of medical decision-making. Certainty is shaped by a constellation of factors, including the caliber and abundance of available evidence, the clinician's expertise, and the patient's unique circumstance. The quality and quantity of available evidence hold immense sway over certainty. When robust

evidence supports a diagnosis or treatment, clinicians are more apt to exercise confidence in their decision-making. However, when evidence is scarce or contradictory, clinicians may grapple with uncertainty.

Clinician experience serves as another cornerstone of certainty. Clinicians with extensive experience are generally more likely to exhibit confidence in their decision-making, having encountered a broader spectrum of patients and conditions. Yet, even seasoned clinicians can err, underscoring the significance of meticulously scrutinizing all available evidence before reaching a verdict. The patient's individual circumstances can also modulate certainty. For instance, if a patient presents with a rare or complex condition, clinicians may face heightened uncertainty regarding the most appropriate course of treatment. Moreover, patient preferences and values can also play a crucial role in decision-making (17).

To quote Sir William Osler, the father of modern medicine who initiated bedside clinical education: "Medicine is a science of uncertainty and an art of probability" (16).

Uncertainty lies at the core of all medical interactions. Medical practice is inherently entwined with uncertainty stemming from a multitude of factors, including technical uncertainties arising from insufficient data, interpersonal uncertainties rooted in the complexities of the physician-patient relationship, and conceptual uncertainties emerging from the application of broad principles to specific scenarios. The inherent generality of scientific knowledge necessitates a nuanced approach to address the unique intricacies of each patient's illness. Consequently, intrinsic uncertainty remains an inextricable component of healthcare. The inherent elusiveness of medical interventions underscores the importance of humility among medical professionals in order to foster a more reciprocal and dialectical approach to the practice of medicine (17).

Role of automation bias

One of the risks that undermines certainty in clinical decision-making is automation bias. In the past, whenever a physician made a judgment about a patient, that judgment was considered definitive. Physicians would evaluate all aspects and take into account sufficient evidence before arriving at a clinical decision, which was deemed adequate. However, with the daily expansion of artificial intelligence capabilities, the opinions provided by AI are increasingly presented to physicians with a sense of authority.

In one study, it was observed that after incorrect responses were given by an AI system regarding an electrocardiogram, physicians' treatment decisions for patients changed significantly (5). The question that arises here is: Can certainty still be leveraged for change, despite awareness of the shortcomings of artificial intelligence? What impact does AI, with all its flaws, have on the automation of patient care and physicians' practices? Can a physician easily alter their opinions about a patient based solely on AI recommendations?

Tolerance of ambiguity leads medicine to be Gentler

In his book *Medical Nihilism*, Jacob Stegenga recommends that doctors adopt a more humble approach in cases such as those mentioned above. Knowing that magic bullets are rare in medicine and that many of the diagnoses and treatments are incorrect, doctors should proceed with more caution. They should prescribe fewer treatments and focus more on disease prevention.

The concept of gentle medicine is not a call for the wholesale abandonment of medical intervention, and there are always a number of effective treatments that should be utilized. Rather, gentle medicine advocates for a more restrained approach to medical intervention, with physicians intervening less frequently, perhaps even much less frequently than they do today. In addition, gentle medicine emphasizes the importance of improving health through lifestyle and societal changes (17).

When we become aware of the problems and limitations of artificial intelligence, it is a mistake to give these achievements more credit than they deserve. Artificial intelligence and large language models can be very helpful in the field of medicine, and their use should definitely be expanded. However, they should always be viewed critically. This is where gentle medicine comes in. Physicians should be aware that these tools can have significant limitations, which can add to uncertainty. As proposed by Stegenga, a more effective approach to improving health would be to focus on restructuring society in ways that address the underlying causes of disease, rather than relying on the "magic bullet" model of medical intervention that has dominated medicine in the past century (17).

In the absence of absolute certainty, clinicians must weigh the risks and benefits of different treatment options and make a decision that they believe is in the best interests of the patient. This may involve making a trade-off between certainty and other factors, such as the patient's preferences and the availability of resources.

In the past, some novel breakthroughs have been put aside just because they did not pass the certainty tests. Having said that, there has always been an acceptable amount of uncertainty when doctors face a problem. There are always errors occurring due to human boundaries, exhaustion, overload, burnout, and ideas simply not coming to mind at the right moment.

It seems that our "schizophrenic friends" (large language models), which can provide significant assistance to doctors, confront us with a new certainty that does not necessarily stem from the old medical traditions. The uncertainty of a doctor's decision-making could be due to a lack of information, and now these language models may also give us incorrect answers as a result of the aforementioned factors.

By reminding practitioners of the inherent uncertainty that underpins medical interventions, the philosophy of medicine fosters a more receptive stance toward these technological advancements. Furthermore, by acknowledging the inherent fallibility of these technologies, the importance of gentle medicine and humility in clinical practice will be reinforced. Physicians must not shy away from embracing AI tools due to their imperfections, but should develop a more accepting attitude toward them by acknowledging their uncertainty. Moreover, by constantly highlighting the imperfections, the philosophy of medicine cultivates a deeper sense of humility among practitioners.

Are we facing a new sort of ambiguity? Is there an acceptable range for AI language models to be uncertain? How will the ethics and legislation deal with this new era? Do the advantages of utilizing such language models outweigh their uncertainty?

Jacques Derrida's concept of the "pharmakon" is a useful lens through which to view the potential impact of AI language models on medicine. A pharmakon is something that can be both beneficial and harmful, depending on how it is used (9). Similarly, AI language models have the potential to improve medical care, but they also pose some risks.

Medical colleges and faculties have a responsibility to ensure that AI language models are used in a responsible and ethical way. They should develop training programs for medical professionals on how to use AI language models effectively and safely. They should also conduct research on the potential risks and benefits of AI language models in medicine.

If we as medical professionals take AI technologies for granted, tech unicorn companies will happily replace our responsibility. This would be a mistake,

as tech companies may not have the same values or priorities as medical professionals. For example, a tech unicorn company may be more interested in profits than in providing high-quality medical care. It is important to remember that AI language models are tools, and like any tool, they can be used for good or for bad. It is up to us to decide how we will use AI language models in medicine. We must use them in a way that is ethical, responsible, and in the best interests of our patients.

Conclusion

In this article, we have argued that AI hallucination and AI drift are two major challenges that we face as we move toward a future where AI models are increasingly used in medicine. These challenges raise a number of ethical questions. We have explored the philosophical and theoretical basis of the uncertainty of AI models in medicine. We have also discussed how we can interface with this new ambiguous era. There are two different aspects of the case that should not be overlooked. First, we must remember that AI flaws will give medical professionals a more humble and gentle approach, which will take place as soon as there is an acceptance in medical settings for these new technological advancements. Second, by acknowledging the uncertain and ambiguous nature of medicine, we will be more open to these new tools. This is where medical humanities plays an important role in answering such ethical questions.

The philosophy of medicine plays a vital part in fostering this awareness, reminding practitioners of the uncertainties that accompany medical interventions. By promoting a balanced perspective, we can harness the benefits of AI while ensuring that ethical considerations and human expertise remain at the forefront of medical practice. As we navigate this complex landscape, it is crucial for both medical professionals and philosophers to collaborate, guiding the responsible and ethical implementation of AI in healthcare.

We started the article with a quote from Aeneid by Virgil: *"Facilis descensus averno: noctes atque dies patet atri ianua ditis; sed revocare gradium superasque evadere ad auras, hoc opus his labor est."* (The gates of hell are open night and day. Smooth the decent and easy is the way, but to return and view the cheerful skies, in this task and mighty labor lies) (7). We strongly support any sort of progression in the medical field that could possibly palliate patient suffering and pain. It is easy to be completely open and receptive to AI, but if we neglect the possible frauds and problems that this

new friend could bring us, we will be faced with new problems and challenges. Under the circumstances, it seems appropriate to practice more gentle medicine. This is where we have to be focused and revise our attempts, as *in this task and mighty labor lies*.

The limitations encountered while writing this article include the rapid growth and daily updates of large language models. Additionally, exploring practical and technical approaches to reduce errors and mistakes in language models should be addressed in further research.

Acknowledgements

The authors are grateful to Dr. Ali Malaeke and Dr.Younes Shokrkhah for their help with the first drafts of the article.

Conflict of interest

There is not any conflict of interest among authors

Funding

This research did not receive any funds.

References:

1. Adamopoulou E, Moussiades L. An overview of chatbot technology. IFIP Int Conf Artif Intell Appl Innov. 2020.
2. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus. 2023;15(2). doi:10.7759/cureus.35179.
3. Aydin Ö. Google Bard generated literature review: metaverse. J AI. 2023;7(1):1-14.
4. Beutel G, Geerits E, Kielstein JT. Artificial hallucination: GPT on LSD? Crit Care. 2023;27(1):148.
5. Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, et al. Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J Electrocardiol. 2018;51(6):S6-S11.
6. Cambria E, Mao R, Chen M, Wang Z, Ho SB. Seven pillars for the future of AI. IEEE Intell Syst. 2023;38(6).
7. Chandon G, Virgil. Stories from the Aeneid. World Pub Co; 1965.
8. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023;6:1169595.

9. Kakoliris G. The "Undecidable" Pharmakon: Derrida's Reading of Plato's Phaedrus. In: The New Yearbook for Phenomenology and Phenomenological Philosophy. Routledge; 2015. p. 231-42.
10. Lee H. The rise of ChatGPT: Exploring its potential in medical education. Anat Sci Educ. 2023.
11. Marcum JA. An introductory philosophy of medicine: Humanizing modern medicine. Vol. 99. Springer Science & Business Media; 2008.
12. Morreel S, Verhoeven V, Mathysen D. Microsoft Bing outperforms five other generative artificial intelligence chatbots in the Antwerp University multiple choice medical license exam. medRxiv. 2023:2023.08.18.23294263.
13. Possati LM. Psychoanalyzing artificial intelligence: The case of Replika. AI Soc. 2023;38(4):1725-38.
14. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber Phys Syst. 2023.
15. Reshamwala A, Mishra D, Pawar P. Review on natural language processing. IRACST Eng Sci Technol Int J. 2013;3(1):113-6.
16. Rysavy M. Evidence-based medicine: a science of uncertainty and an art of probability. AMA J Ethics. 2013;15(1):4-8.
17. Stegenga J. Medical nihilism. 1st ed. Oxford University Press; 2018.
18. Sukel M, Rudinac S, Worring M. GIGO, Garbage In, Garbage Out: An Urban Garbage Classification Dataset. Int Conf Multimed Model. 2023.
19. Thapa S, Adhikari S. ChatGPT, bard, and large language models for biomedical research: opportunities and pitfalls. Ann Biomed Eng. 2023:1-5.
20. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40.