

Viewpoint

Large Language Models and User Trust: Consequence of Self-Referential Learning Loop and the Deskilling of Health Care Professionals

Avishek Choudhury, PhD; Zaira Chaudhry, MD

Industrial and Management Systems Engineering, West Virginia University, Morgantown, WV, United States

Corresponding Author:

Avishek Choudhury, PhD

Industrial and Management Systems Engineering

West Virginia University

321 Engineering Sciences Bldg

1306 Evansdale Drive

Morgantown, WV, 26506

United States

Phone: 1 5156080777

Email: avishek.choudhury@mail.wvu.edu

Abstract

As the health care industry increasingly embraces large language models (LLMs), understanding the consequence of this integration becomes crucial for maximizing benefits while mitigating potential pitfalls. This paper explores the evolving relationship among clinician trust in LLMs, the transition of data sources from predominantly human-generated to artificial intelligence (AI)-generated content, and the subsequent impact on the performance of LLMs and clinician competence. One of the primary concerns identified in this paper is the LLMs' self-referential learning loops, where AI-generated content feeds into the learning algorithms, threatening the diversity of the data pool, potentially entrenching biases, and reducing the efficacy of LLMs. While theoretical at this stage, this feedback loop poses a significant challenge as the integration of LLMs in health care deepens, emphasizing the need for proactive dialogue and strategic measures to ensure the safe and effective use of LLM technology. Another key takeaway from our investigation is the role of user expertise and the necessity for a discerning approach to trusting and validating LLM outputs. The paper highlights how expert users, particularly clinicians, can leverage LLMs to enhance productivity by off-loading routine tasks while maintaining a critical oversight to identify and correct potential inaccuracies in AI-generated content. This balance of trust and skepticism is vital for ensuring that LLMs augment rather than undermine the quality of patient care. We also discuss the risks associated with the deskilling of health care professionals. Frequent reliance on LLMs for critical tasks could result in a decline in health care providers' diagnostic and thinking skills, particularly affecting the training and development of future professionals. The legal and ethical considerations surrounding the deployment of LLMs in health care are also examined. We discuss the medicolegal challenges, including liability in cases of erroneous diagnoses or treatment advice generated by LLMs. The paper references recent legislative efforts, such as The Algorithmic Accountability Act of 2023, as crucial steps toward establishing a framework for the ethical and responsible use of AI-based technologies in health care. In conclusion, this paper advocates for a strategic approach to integrating LLMs into health care. By emphasizing the importance of maintaining clinician expertise, fostering critical engagement with LLM outputs, and navigating the legal and ethical landscape, we can ensure that LLMs serve as valuable tools in enhancing patient care and supporting health care professionals. This approach addresses the immediate challenges posed by integrating LLMs and sets a foundation for their maintainable and responsible use in the future.

(*J Med Internet Res* 2024;26:e56764) doi: [10.2196/56764](https://doi.org/10.2196/56764)

KEYWORDS

trust; ChatGPT; human factors; healthcare; LLMs; large language models; LLM user trust; AI accountability; artificial intelligence; AI technology; technologies; effectiveness; policy; medical student; medical students; risk factor; quality of care; healthcare professional; healthcare professionals; human element

Introduction

Overview

Integration of existing artificial intelligence (AI) models into health care—a field where the trust in AI is crucial due to the significant impact of decision-making—is still a work in progress [1]. At the same time, efforts to develop standardized protocols for the deployment of AI in health care are underway, yet they have not reached a point of completion [2]. This endeavor is critical for ensuring AI's safe and effective use in health care settings. Additionally, the challenge of evaluating AI in health care is exacerbated by a lack of comprehensive and standardized metrics [3]. This void is something that researchers and policymakers are actively working to address by creating robust evaluation frameworks that could be applied universally. The regulatory landscape has been focusing on policies around ethical considerations, data privacy, transparency, and patient safety, alongside frameworks that hold AI systems and their developers accountable for the outcomes of their use in patient care [1].

Advent of Generative AI—Large Language Models in Health Care

Despite these ongoing challenges and developments, generative AI like large language models (LLM) is already being deployed in the public sphere [4,5], used by health care workers, researchers, and the public for a variety of health care-related tasks. Although LLMs have shown promise in medical assessments [6-10], scientific writing, eHealth care, and patient classification [11-13], their integration marks a shift in paradigm introducing new AI complexities [14-17]. Its rapid and early adoption highlights the critical need for continued discourse ensuring the safe and effective integration of LLM into health care. Additionally, LLM characters such as—stochasticity, emergent indeterminacy, and lack of consciousness—reinforces the need for cautiousness.

One fundamental aspect of LLMs that prompts special attention is their stochastic paradigm, which means that these models operate based on probabilities and randomness, allowing the model to generate varied outputs for a given input. It exhibits a level of indeterminacy and unpredictability. LLMs can produce different responses under seemingly similar conditions, complicating their reliability. Such LLM behaviors can lead to unexpected results, which, while sometimes beneficial in generating creative solutions or insights, can also pose risks when applied to critical domains like health care, where accuracy and predictability are paramount.

Another critical risk characteristic of LLM is the lack of inherent understanding of the context they parse and generate. Despite their ability to produce human-like text, LLMs do not possess consciousness, comprehension, or the ability to discern the truthfulness of their outputs. In other words, LLMs might generate plausible but incorrect content, presenting significant challenges in contexts where the veracity and relevance of information are critical [18,19].

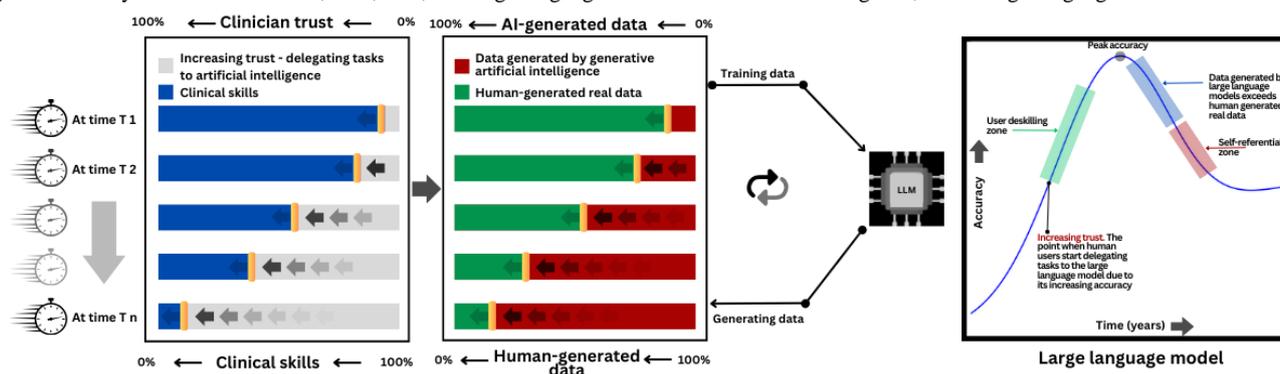
Approaching AI integration in health care with a critical mindset is important. It is crucial for users to have a clear understanding of a technology's actual performance, distinguishing it from the exaggerated expectations set by media hype. These risks underscore the importance of asking the question: are we and our health care system ready to integrate LLMs? If yes, is there a policy in place explicitly stating in what capacity it could be used to reduce clinical workload before its dissemination?

Objective

In this paper, we conceptually investigate the dynamics between clinicians' growing trust in LLMs, the evolving sources of training data, and the resultant implications for both clinician competency and LLM performance over time. Our discussion highlights a potential feedback loop where LLMs, increasingly trained on narrower data sets dominated by their own outputs, may experience a decline in output quality coinciding with a reduction in user skills. While these phenomena are not yet fully realized, they represent anticipated challenges that coincide with the deeper integration of LLMs into the health care domain. We call for preemptive, focused dialogues concerning the integration of LLMs in medical settings, underscoring the importance of maintaining patient safety and the standard of care.

Presently, LLMs are developing at an accelerated pace, heavily reliant on human-generated data sets that are integral to their accuracy and the consequent trust placed in them, particularly in the health care sector. This burgeoning dependency, although seemingly beneficial in terms of efficiency and productivity, may lead to an unintended erosion of clinician skills due to the habitual delegation of tasks to AI—as noted in the academic context [20,21]. This trend raises the possibility of an overreliance on LLM outputs, potentially diminishing the variety and depth of human insights within these models. The risk is a self-perpetuating cycle where LLMs, learning mostly from their creations, could see a degradation in their effectiveness and a narrowing of the breadth of human knowledge they were designed to emulate. Such an outcome would be counterproductive, possibly leading to a decline in both LLM effectiveness and human expertise.

Figure 1 illustrates our core arguments. The first panel reveals a timeline that shows an inverse correlation between clinicians' escalating trust in AI and the preservation of clinical skills over successive time points (T1 to Tn), signaling an increase in AI reliance and a decrease in skill retention. The middle panel demonstrates the shift in training data for LLMs from predominantly human-generated to a growing proportion of AI-generated data, which in turn affects LLM performance and contributes to the feedback loop. The final panel plots LLM accuracy against time, displaying an initial increase as LLMs leverage a mix of data sources. However, upon reaching a tipping point—marked as the self-referential zone—accuracy declines in tandem with the onset of the user deskilling zone, emphasizing the dilemma of increased AI reliance degrading user capabilities. We underscore the need for strategic measures to address these impending challenges in the health care sector.

Figure 1. The dynamics of user skills, trust, data, and large language models. AI: artificial intelligence; LLM: large language model.

User Expertise and Trust in LLMs

Overview

User trust in LLMs is deeply intertwined with the individual's subject matter expertise and their willingness to engage critically with AI outcomes. Expert users, with a robust understanding of their domain, are more likely to approach LLMs with a discerning mindset and preparedness to review and validate its suggestions. Thus, trust in LLMs can be seen as a spectrum influenced by the user's expertise, and the effort they are willing to invest in ensuring the accuracy of the outcomes.

User Expertise: Ability to Detect Errors in LLMs

The use of LLMs presents a range of possibilities and challenges that vary depending on the user's expertise and intent, delineating into 2 primary user categories—subject matter experts and those seeking assistance due to a lack of knowledge.

Subject matter experts (doctors) may use LLMs to handle routine, time-consuming tasks, enabling them to allocate more time to complex or urgent issues like seeking a second opinion on complex medical diagnoses, or patient triage. They have the advantage of being able to critically evaluate the LLM's output, verify its accuracy (deviation from clinical standards), and make necessary corrections. The expertise of such users acts as a safeguard against potential errors, ensuring that the AI's assistance enhances productivity without introducing risk.

On the other hand, individuals who turn to LLMs due to a lack of expertise in a particular area face a different set of challenges [22]. The ability of LLMs to generate fake but persuasive responses further exacerbates the risks making users vulnerable to accepting erroneous information as fact [23]. For instance, a general practitioner faced with a dermatological case, such as an atypical presentation of psoriasis, can use an LLM to access detailed diagnostic criteria and treatment protocols. This capability can significantly assist in the management of the patient, particularly when the LLM's suggestions are accurate and relevant. However, the inherent risk of LLMs generating incorrect suggestions cannot be overlooked. Such inaccuracies pose a heightened risk to patient safety, especially in scenarios where the clinician may lack the specialized dermatological knowledge required to critically evaluate the validity of the LLM's output [24]—constituting an environment where trust becomes critical.

The crux of the problem lies in the user's ability to verify the accuracy and relevance of the AI-generated content. However, the pivotal consideration here is whether the verification of LLM outcomes by health care staff negates the purported reduction in workload. If health care professionals are required to meticulously check each AI-generated output for accuracy, the time saved through automation may be offset by the time spent on verification. Maintaining the balance between productivity and accuracy is pivotal. For instance, LLMs can analyze vast data sets to identify patterns or treatment outcomes that may not be immediately apparent to human clinicians, thereby offering insights that can lead to more accurate diagnoses and personalized treatment plans. This capability, even if it requires additional time for verification of AI-generated recommendations, may be deemed a worthy trade-off for reducing long-term health care costs and improving care quality. However, this trade-off must be carefully managed to ensure that the pursuit of improved health outcomes does not lead to unsustainable decreases in productivity. Excessive time spent verifying AI recommendations could strain health care resources, leading to longer patient wait times and potentially overburdening health care staff.

To navigate this trade-off, health care systems might adopt strategies such as targeted use of LLMs in high-impact areas where they are most likely to enhance outcomes and the development of systems that prioritize clarity and actionability in their recommendations to minimize verification time. By carefully weighing the benefits of improved patient outcomes against the costs in terms of productivity, health care providers can make informed decisions about how best to integrate LLMs into their practices, ensuring that these technologies serve to enhance rather than hinder the delivery of patient care.

User Trust: Willingness to Review LLM Output

Trust in user engagement with LLMs, particularly in health care, is a multifaceted construct influenced by sociotechnical and psychological factors. We acknowledge that user trust in LLMs, in health care, can substantially depend on the context. Depending on the stakes (risk) the level of trust required may differ; for instance, LLMs used for diagnosis and treatment recommendations necessitate a higher trust level compared to applications for patient note summarization. Additionally, the degree of autonomy granted to the LLMs, and the extent of clinical oversight are crucial determinants of trust.

Clinicians bring their own norms and expectations to the evaluation of trust in these systems, further complicating the landscape. Individual and cultural perspectives on risk tolerance and acceptance also play pivotal roles. Together, these factors create a complex environment where trust in LLMs is dynamic, varying according to the specific context of use and the interplay of diverse elements. In this section, we focus on user willingness to scrutinize LLM output as a precursor to trust.

A user may have the ability and necessary expertise but may not be willing to review LLM-generated outcomes due to factors including prior trust in the technology or biases. A doctor with high trust (blind trust) in the LLM, might be more inclined to accept its suggestion without extensive further verification [25,26], exhibiting automation bias [27]. Automation bias, particularly in the context of clinicians' interactions with LLMs can manifest when clinicians exhibit an undue level of trust in the systems, based on past experiences of accuracy and reliability.

Blind trust in LLMs can introduce 2 critical cognitive biases, precautionary [28] and confirmation bias [29], both of which alter clinician behavior in the presence of agreement or disagreement between human judgment and LLM outputs. When LLM recommendations align with a clinician's initial diagnosis or treatment plan (agreement), confirmation bias can be reinforced. Clinicians may overlook or undervalue subsequent information that contradicts the LLM-supported decision, even if this new information is critical to patient care. This confirmation bias can lead to a narrowed diagnostic vision, where alternative diagnoses or treatments are not sufficiently considered. Conversely, in cases where there is a disagreement, precautionary bias can occur. The clinician, having developed a reliance on the LLM due to positive past experiences, might doubt their own expertise and perceive LLM to be the safer alternative for decision-making. Such problems associated with blind trust might persist unchallenged until a point of failure or harm, which can have serious implications in health care.

Future Risk Considerations

Overview

As we delve deeper into the dynamics between technology and human expertise, the concepts of the LLM Paradox of Self-Referential Loop and the Risk of Deskilling emerge as pivotal to our discourse. As Figure 1 illustrates the projected trajectory of clinician reliance on LLMs but also hints at the potentially cyclic nature of knowledge and skills within the health care industry. Concurrently, the risk of deskilling looms over the horizon, particularly for upcoming generations of health care professionals who might become overly reliant on LLMs, possibly at the expense of their diagnostic acumen and critical thinking abilities. This section explores these challenges and the strategies needed to mitigate them. Additionally, this section discusses the LLM accountability concern.

LLM Paradox of Self-Referential Loop (Learning From Itself)

In a scenario where LLMs become widely adopted in the health care industry for tasks like paper writing, educational material creation, clinical text summarization, and risk identification,

the possibility of a self-referential loop does emerge as a significant concern. This paradox occurs when AI-generated human-like content becomes so widespread that the AI begins to reference its own generated content, potentially leading to an echo chamber effect where original, human-generated insights become diluted or harder to distinguish from AI-generated content. While this problem of a self-referential loop in AI-generated content, particularly in the health care industry, has not yet materialized, it represents a likely challenge as generative AI continues to proliferate. The consequence of a self-referential loop in LLMs can lead to several problematic outcomes, including the propagation of biases [30], increased homogeneity in generated data, and ultimately, hindered performance. AI systems learn from the data they are fed, and if these data include biases, the AI is likely to replicate and even amplify these biases in its outputs [31]. In a self-referential loop, the problem becomes compounded. As the AI references its own biased outputs to generate new content, these biases can become more entrenched, making them harder to identify and correct.

The issue of self-referential loops and the potential degradation of information quality are indeed significant concerns; however, when these LLMs, such as the Medical Pathways Language Model (Med-PaLM) [32], are specifically fine-tuned and tailored for health care applications, the severity of these issues can be mitigated through stringent quality assurance measures. This approach reduces the risk associated with the indiscriminate use of a broader corpus that may contain inaccuracies, outdated information, or irrelevant content. Despite these precautions, the risk of self-referential loops in health care contexts can shift toward a different concern, the reinforcement and entrenchment of specific clinical approaches and schools of thought. This occurs as a reflection of the biases present in the curated data sets, which are inherently influenced by the prevailing medical practices, research focus, and therapeutic approaches at the time of data collection.

Addressing this challenge requires a nuanced approach to developing and integrating LLMs technologies into societal frameworks. It involves fostering a symbiotic relationship between human intellect and LLM capabilities, ensuring that AI serves as a tool for augmenting human intellect rather than replacing it. Strategies for maintaining the diversity and quality of training data, including the deliberate inclusion of varied and novel human-generated content, will be critical.

Risk of Deskilling

As individuals come to rely more on LLMs for routine tasks, such as the synthesis of patient information or the interpretation of medical data, there is a possibility that their skills in these critical areas may diminish over time due to reduced practice [33]. This situation is compounded by the AI's ability to quickly furnish answers to medical inquiries, which might decrease the motivation for in-depth research and learning, consequently affecting the professionals' knowledge depth and critical thinking capabilities.

It is crucial to note that the discussion here does not assert that LLMs will definitively lead to the deskilling of current practitioners in the health care sector. These professionals have

developed their expertise through extensive experience and rigorous academic training, establishing a solid foundation that is not readily compromised by the integration of AI tools. Instead, the concern is more pronounced for the next generation of health care professionals, particularly medical students who might increasingly use AI for educational tasks and learning activities where over delegating tasks to AI could attenuate the development of critical analytical skills and a comprehensive understanding of medical concepts, traditionally cultivated through deep engagement with the material [33,34]. The critical question emerges “will the ease of generating content with AI stifle the development of creativity and critical thinking in younger generations accustomed to technology providing immediate solutions?”

If future generations of clinicians grow accustomed to AI doing the bulk of diagnostic review and analysis, there is a risk that their own diagnostic skills might not develop as fully. More critically, should they be required to review patient charts manually—due to AI failures—they may find the task daunting, or lack the detailed insight that manual review processes help to cultivate. The crux of the issue lies in ensuring that reliance on technology should not come at the expense of fundamental skills and knowledge. The challenge is to ensure that the deployment of AI technologies complements human abilities without diminishing the need for critical thinking, reasoning, and creativity.

What is needed is to adapt to the paradigm shift—failing to do so can adversely impact health care industry. A dual focus on harnessing AI capabilities while enhancing unique human skills is pivotal for advancing patient care in the modern medical landscape. The advent of human-AI collaboration in health care prompts a shift in the skill set emphasis within medical disciplines. The transformation accentuates the value of unique human skills—such as problem-solving, critical thinking, creativity, and fostering patient rapport—over traditional reliance on memory and knowledge base tasks. As LLMs undertake roles in diagnostic assistance, literature synthesis, and treatment optimization, the medical profession should evolve to leverage AI for data-driven insights while prioritizing human-centric skills for patient care. The paradigm shift underscores the growing importance of critical engagement with AI outputs, necessitating medical professionals to adeptly interpret and apply AI-generated information within the complex context of individual patient needs.

LLM Accountability

Overview

The integration of LLMs in health care introduces medicolegal challenges concerning the allocation and apportionment of liability for outcomes, particularly in instances of negligent diagnoses and treatment. The complexity arises from the interaction between clinicians, health care institutions, and AI providers, each contributing differently to the health care delivery process.

Legal Framework and Liability Allocation

In the legal domain, traditional frameworks for medical liability often center on direct human actions, with established principles

guiding negligence and malpractice claims. The introduction of LLMs used for diagnostic support or task delegation complicates these frameworks. Clinicians, operating at the interface of LLM recommendations and patient care, are generally seen as the final decision-makers, thus bearing the primary moral and legal responsibility for the outcomes of those decisions. This perspective is grounded in the principle that clinicians must integrate LLM outputs into a broader clinical judgment context, considering patient-specific factors and adhering to professional standards.

Shared Liability and AI Providers

However, the role of LLM providers in developing, deploying, and maintaining LLMs introduces questions about shared liability, especially when system errors or deficiencies contribute to adverse outcomes. Determining the extent of LLM provider liability hinges on factors such as the accuracy of the LLM’s training data, transparency regarding the system’s capabilities and limitations, and the adequacy of user training and support provided.

Institutional Responsibility

Health care institutions also play a critical role in mediating the use of LLMs, responsible for ensuring that these systems are integrated into clinical workflows in a manner that upholds patient safety and complies with regulatory standards. Institutional policies and practices, including the selection of AI tools, clinician training, and oversight mechanisms, are pivotal in mitigating risks associated with LLM use.

Algorithmic Accountability Act of 2023

The Algorithmic Accountability Act of 2023 and Artificial Intelligence Accountability Act [35,36] represent a critical legislative step toward ensuring the responsible use of algorithms. The act calls for the creation of standardized procedures and assessment frameworks to evaluate the effectiveness and consequences of these systems, reflecting an understanding of the complex ethical and regulatory challenges posed by AI in decision-making processes, particularly in health care. The act is in dialogue with the wider conversation on the ethics of AI, advocating for an approach that emphasizes response-ability—the capacity to respond ethically to the challenges posed by algorithmic decision-making. This perspective is crucial for developing impact assessments and frameworks aimed at promoting fairness and preventing discriminatory practices within algorithmic systems.

The implications of this act on the integration of LLMs in health care are profound and ensuring transparency in LLM can further enhance trust in the system. Transparency can allow clinicians to verify errors and review outputs effectively. For example, an LLM providing a diagnostic suggestion would detail the medical literature and patient data informing its analysis, enhancing clinician trust by making the AI’s reasoning processes visible and understandable. This transparency combats algorithmic deference by encouraging health care professionals to critically assess LLM outputs against their expertise and patient-specific contexts. Moreover, transparency reduces the perceived infallibility of LLMs by highlighting their reliance on input data

quality and inherent limitations, promoting a balanced use of LLMs as supportive tools in patient care.

Conclusions

It's important to acknowledge that the performance of LLMs like ChatGPT (OpenAI) today does not guarantee their performance tomorrow. LLMs have the potential to be a substantial boon to the health care industry, offering to streamline workflows, enhance the accuracy of patient data processing, and even support diagnostic and treatment planning processes. Its value, however, is contingent upon a systematic and informed integration into health care systems. Recognizing that LLMs, like any technology, is fallible is crucial to its successful adoption. Its performance is temporal and will change as new data are fed to its algorithm. This acknowledgment underpins the necessity for robust oversight mechanisms,

ongoing evaluation of AI-driven outputs for accuracy and relevance, and clear guidelines on its role as an assistive tool rather than a stand-alone decision-maker.

A thoughtful, deliberate approach to integrating generative AI into health care can mitigate risks associated with overreliance and deskilling, ensuring that it complements rather than compromises the quality of care. By leveraging AI's strengths and compensating for its limitations through human oversight, health care can harness the benefits of this technology to improve outcomes, enhance patient care, and support health care professionals in their vital work. Thus, the path forward involves embracing generative AI's potential while remaining vigilant about its limitations, ensuring that its integration enhances rather than diminishes the human element in health care.

Acknowledgments

This study is not funded by any internal or external agency.

Conflicts of Interest

None declared.

References

1. Harris LA. Artificial intelligence: background, selected issues, and policy considerations. Congressional Research Service. 2021. URL: <https://crsreports.congress.gov/product/pdf/R/R46795> [accessed 2024-04-02]
2. DoC. AI accountability policy request for comment. National Archives. 2023. URL: <https://www.federalregister.gov/documents/2023/04/13/2023-07776/ai-accountability-policy-request-for-comment> [accessed 2024-04-02]
3. Shinnars L, Aggar C, Grace S, Smith S. Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: an integrative review. *Health Informatics J*. 2020;26(2):1225-1236. [FREE Full text] [doi: [10.1177/1460458219874641](https://doi.org/10.1177/1460458219874641)] [Medline: [31566454](https://pubmed.ncbi.nlm.nih.gov/31566454/)]
4. Hegde N, Vardhan M, Nathani D, Rosenzweig E, Speed C, Karthikesalingam A, et al. Infusing behavior science into large language models for activity coaching. *PLOS Digit Health*. Apr 2024;3(4):e0000431-e0000413. [FREE Full text] [doi: [10.1371/journal.pdig.0000431](https://doi.org/10.1371/journal.pdig.0000431)] [Medline: [38564502](https://pubmed.ncbi.nlm.nih.gov/38564502/)]
5. Firaina R, Sulisworo D. Exploring the usage of ChatGPT in higher education: frequency and impact on productivity. *Bul Edukasi Indones*. 2023;2(01):39-46. [doi: [10.56741/bei.v2i01.310](https://doi.org/10.56741/bei.v2i01.310)]
6. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. *Clin Med (Lond)*. 2023;23(3):278-279. [FREE Full text] [doi: [10.7861/clinmed.2023-0078](https://doi.org/10.7861/clinmed.2023-0078)] [Medline: [37085182](https://pubmed.ncbi.nlm.nih.gov/37085182/)]
7. Hung YC, Chaker SC, Sigel M, Saad M, Slater ED. Comparison of patient education materials generated by chat generative pre-trained transformer versus experts: an innovative way to increase readability of patient education materials. *Ann Plast Surg*. 2023;91(4):409-412. [doi: [10.1097/SAP.0000000000003634](https://doi.org/10.1097/SAP.0000000000003634)] [Medline: [37713147](https://pubmed.ncbi.nlm.nih.gov/37713147/)]
8. Galido PV, Butala S, Chakerian M, Agustines D. A case study demonstrating applications of ChatGPT in the clinical management of treatment-resistant schizophrenia. *Cureus*. 2023;15(4):e38166. [FREE Full text] [doi: [10.7759/cureus.38166](https://doi.org/10.7759/cureus.38166)] [Medline: [37252576](https://pubmed.ncbi.nlm.nih.gov/37252576/)]
9. Morrison M, Nobles V, Johnson-Agbakwu CE, Bailey C, Liu L. Classifying refugee status using common features in EMR. *Chem Biodivers*. 2022;19(10):e202200651. [doi: [10.1002/cbdv.202200651](https://doi.org/10.1002/cbdv.202200651)] [Medline: [36050919](https://pubmed.ncbi.nlm.nih.gov/36050919/)]
10. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721-732. [FREE Full text] [doi: [10.3350/cmh.2023.0089](https://doi.org/10.3350/cmh.2023.0089)] [Medline: [36946005](https://pubmed.ncbi.nlm.nih.gov/36946005/)]
11. Chen TJ. ChatGPT and other artificial intelligence applications speed up scientific writing. *J Chin Med Assoc*. 2023;86(4):351-353. [FREE Full text] [doi: [10.1097/JCMA.0000000000000900](https://doi.org/10.1097/JCMA.0000000000000900)] [Medline: [36791246](https://pubmed.ncbi.nlm.nih.gov/36791246/)]
12. Malamas N, Papangelou K, Symeonidis AL. Upon improving the performance of localized healthcare virtual assistants. *Healthcare (Basel)*. 2022;10(1):99. [FREE Full text] [doi: [10.3390/healthcare10010099](https://doi.org/10.3390/healthcare10010099)] [Medline: [35052263](https://pubmed.ncbi.nlm.nih.gov/35052263/)]
13. Wang DQ, Feng LY, Ye JG, Zou JG, Zheng YF. Accelerating the integration of ChatGPT and other large - scale AI models into biomedical research and healthcare. *MedComm—Future Medicine*. 2023;2(2):1-28. [FREE Full text] [doi: [10.1002/mef2.43](https://doi.org/10.1002/mef2.43)]
14. Hoffmann J, Borgeaud S, Mensch A, Buchatskaya E, Cai T, Rutherford E, et al. Training compute-optimal large language models. ArXiv. preprint posted online on March 29, 2022. [FREE Full text] [doi: [10.48550/arxiv.2203.15556](https://doi.org/10.48550/arxiv.2203.15556)]

15. Seth I, Kenney PS, Bulloch G, Hunter-Smith DJ, Thomsen JB, Rozen WM. Artificial or augmented authorship? A conversation with a chatbot on base of thumb arthritis. *Plast Reconstr Surg Glob Open*. 2023;11(5):e4999. [FREE Full text] [doi: [10.1097/GOX.0000000000004999](https://doi.org/10.1097/GOX.0000000000004999)] [Medline: [37250832](https://pubmed.ncbi.nlm.nih.gov/37250832/)]
16. Yeung JA, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, et al. AI chatbots not yet ready for clinical use. *Front Digit Health*. 2023;5:1161098. [FREE Full text] [doi: [10.3389/fdgth.2023.1161098](https://doi.org/10.3389/fdgth.2023.1161098)] [Medline: [37122812](https://pubmed.ncbi.nlm.nih.gov/37122812/)]
17. Cloesmeijer ME, Janssen A, Koopman SF, Cnossen MH, Mathôt RAA, SYMPHONY consortium. ChatGPT in pharmacometrics? Potential opportunities and limitations. *Br J Clin Pharmacol*. Jan 2024;90(1):360-365. [doi: [10.1111/bcp.15895](https://doi.org/10.1111/bcp.15895)] [Medline: [37621112](https://pubmed.ncbi.nlm.nih.gov/37621112/)]
18. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1166120. [FREE Full text] [doi: [10.3389/fpubh.2023.1166120](https://doi.org/10.3389/fpubh.2023.1166120)] [Medline: [37181697](https://pubmed.ncbi.nlm.nih.gov/37181697/)]
19. Gravel J, D'Amours-Gravel M, Osmanliu E. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. *MCP: Digital Health*. 2023;1(3):226-234. [FREE Full text] [doi: [10.1016/j.mcpdig.2023.05.004](https://doi.org/10.1016/j.mcpdig.2023.05.004)]
20. Sison AJG, Daza MT, Gozalo-Brizuela R, Garrido-Merchán EC. ChatGPT: more than a "weapon of mass deception" ethical challenges and responses from the Human-Centered Artificial Intelligence (HCAI) perspective. *Int J Hum Comput*. 2023.:1-20. [doi: [10.1080/10447318.2023.2225931](https://doi.org/10.1080/10447318.2023.2225931)]
21. Birenbaum M. The chatbots' challenge to education: disruption or destruction? *Educ Sci (Basel)*. 2023;13(7):711. [FREE Full text] [doi: [10.3390/educsci13070711](https://doi.org/10.3390/educsci13070711)]
22. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int J Inf Manag*. 2023;71:102642. [FREE Full text] [doi: [10.1016/j.ijinfomgt.2023.102642](https://doi.org/10.1016/j.ijinfomgt.2023.102642)]
23. Shahsavari Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. *JMIR Hum Factors*. 2023;10:e47564. [FREE Full text] [doi: [10.2196/47564](https://doi.org/10.2196/47564)] [Medline: [37195756](https://pubmed.ncbi.nlm.nih.gov/37195756/)]
24. Dempere J, Modugu K, Hesham A, Ramasamy LK. The impact of ChatGPT on higher education. *Front Educ*. 2023;8:1206936. [FREE Full text] [doi: [10.3389/feduc.2023.1206936](https://doi.org/10.3389/feduc.2023.1206936)]
25. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *J Med Internet Res*. 2023;25:e47184. [FREE Full text] [doi: [10.2196/47184](https://doi.org/10.2196/47184)] [Medline: [37314848](https://pubmed.ncbi.nlm.nih.gov/37314848/)]
26. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res*. 2020;22(6):e15154. [FREE Full text] [doi: [10.2196/15154](https://doi.org/10.2196/15154)] [Medline: [32558657](https://pubmed.ncbi.nlm.nih.gov/32558657/)]
27. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121-127. [FREE Full text] [doi: [10.1136/amiajnl-2011-000089](https://doi.org/10.1136/amiajnl-2011-000089)] [Medline: [21685142](https://pubmed.ncbi.nlm.nih.gov/21685142/)]
28. Vlassov VV. Precautionary bias. *Eur J Public Health*. 2017;27(3):389. [FREE Full text] [doi: [10.1093/eurpub/ckx013](https://doi.org/10.1093/eurpub/ckx013)] [Medline: [28549138](https://pubmed.ncbi.nlm.nih.gov/28549138/)]
29. Oswald ME, Grosjean S. Confirmation bias. In: *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Hove, East Sussex, United Kingdom. Psychology Press; 2004;79-96.
30. Meyer JG, Urbanowicz RJ, Martin PCN, O'Connor K, Li R, Peng PC, et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Min*. 2023;16(1):20. [FREE Full text] [doi: [10.1186/s13040-023-00339-9](https://doi.org/10.1186/s13040-023-00339-9)] [Medline: [37443040](https://pubmed.ncbi.nlm.nih.gov/37443040/)]
31. Nazir A, Wang Z. A comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. *Meta Radiol*. 2023;1(2):100022. [FREE Full text] [doi: [10.1016/j.metrad.2023.100022](https://doi.org/10.1016/j.metrad.2023.100022)] [Medline: [37901715](https://pubmed.ncbi.nlm.nih.gov/37901715/)]
32. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. [FREE Full text] [doi: [10.1038/s41586-023-06291-2](https://doi.org/10.1038/s41586-023-06291-2)] [Medline: [37438534](https://pubmed.ncbi.nlm.nih.gov/37438534/)]
33. Lam K. ChatGPT for low- and middle-income countries: a Greek gift? *Lancet Reg Health West Pac*. 2023;41:100906. [FREE Full text] [doi: [10.1016/j.lanwpc.2023.100906](https://doi.org/10.1016/j.lanwpc.2023.100906)] [Medline: [37745974](https://pubmed.ncbi.nlm.nih.gov/37745974/)]
34. Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. *Educ Sci (Basel)*. 2023;13(4):410. [FREE Full text] [doi: [10.3390/educsci13040410](https://doi.org/10.3390/educsci13040410)]
35. H.R.3369—118th congress (2023-2024). Congress. 2023. URL: <https://www.congress.gov/bill/118th-congress/house-bill/3369> [accessed 2024-04-02]
36. Algorithmic accountability act of 2023. Office of U.S. Senator Ron Wyden. 2023. URL: <https://www.congress.gov/bill/118th-congress/senate-bill/2892/text> [accessed 2024-04-02]

Abbreviations

AI: artificial intelligence

LLM: large language model

Med-PaLM: Medical Pathways Language Model

Edited by T de Azevedo Cardoso, G Eysenbach; submitted 25.01.24; peer-reviewed by M Saremi, D Hua; comments to author 08.03.24; revised version received 12.03.24; accepted 20.03.24; published 25.04.24

Please cite as:

Choudhury A, Chaudhry Z

Large Language Models and User Trust: Consequence of Self-Referential Learning Loop and the Deskillling of Health Care Professionals
J Med Internet Res 2024;26:e56764

URL: <https://www.jmir.org/2024/1/e56764>

doi: [10.2196/56764](https://doi.org/10.2196/56764)

PMID: [38662419](https://pubmed.ncbi.nlm.nih.gov/38662419/)

©Avishek Choudhury, Zaira Chaudhry. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 25.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.