

Implications of conscious AI in primary healthcare

Dorsai Ranjbari,¹ Samira Abbasgholizadeh Rahimi ^{2,3,4}

To cite: Ranjbari D, Abbasgholizadeh Rahimi S. Implications of conscious AI in primary healthcare. *Fam Med Community Health* 2024;**12**:e002625. doi:10.1136/fmch-2023-002625

Received 31 October 2023
Accepted 29 February 2024

ABSTRACT

The conversation about consciousness of artificial intelligence (AI) is an ongoing topic since 1950s. Despite the numerous applications of AI identified in healthcare and primary healthcare, little is known about how a conscious AI would reshape its use in this domain. While there is a wide range of ideas as to whether AI can or cannot possess consciousness, a prevailing theme in all arguments is uncertainty. Given this uncertainty and the high stakes associated with the use of AI in primary healthcare, it is imperative to be prepared for all scenarios including conscious AI systems being used for medical diagnosis, shared decision-making and resource management in the future. This commentary serves as an overview of some of the pertinent evidence supporting the use of AI in primary healthcare and proposes ideas as to how consciousnesses of AI can support or further complicate these applications. Given the scarcity of evidence on the association between consciousness of AI and its current state of use in primary healthcare, our commentary identifies some directions for future research in this area including assessing patients', healthcare workers' and policy-makers' attitudes towards consciousness of AI systems in primary healthcare settings.

against the idea that machines cannot be conscious.² Although experts currently have the consensus that none of the existing AI systems have consciousness, with incremental improvements of the existing AI, we might be facing systems that would make us doubt their consciousness.³

Turing's test has been pivotal for discussions around AI's consciousness, but it is not certain whether passing this test is enough to consider an AI system to be conscious. Searle's Chinese room argument, for example, states that an AI system's capacity for producing meaningful outputs does not necessitate the involvement of consciousness understanding in the internal processes of that system.⁴

Presently, there is a wide range of arguments about AI's consciousness. There are arguments stating that consciousness is reserved for creatures who have the same specific causal biochemical neural structure as many animals do.³ On the contrary, there are philosophers and scientists who see consciousness as an 'internal model' of attention and suppose that it is programmable into artificial intelligence the same way our biological neural structure codes for human consciousness.⁵ There currently are at least 22 theories of consciousness which attempt to correlate consciousness with its neurobiological basis.⁶ A few standouts among these existing theories are global neuronal workspace theory,⁷ higher order theory⁸ and information integration theory of consciousness.⁹

The global workspace theory claims that consciousness is created when a sudden and exclusive activation of a specific subset of neurons associated with an experience, such as a specific perception, broadcast that experience and make it available to local processors such as memory, attention or verbal reports.^{6,7} The higher order theory, on the other hand, postulates that an entity can only be conscious when it is, at least to some extent, aware of its internal processes through meta representations of lower-level representations—such as a sensation—in higher level processing parts such as the prefrontal cortex



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹McGill University Faculty of Medicine and Health Sciences, Montreal, Quebec, Canada

²Family Medicine, Faculty of Medicine and Health Sciences and Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Quebec, Canada

³Mila - Quebec AI Institute, Montreal, Quebec, Canada

⁴Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

Correspondence to

Dr Samira Abbasgholizadeh Rahimi;
samira.rahimi@mcgill.ca

INTRODUCTION

Do artificial intelligence (AI) enabled systems have consciousness? The definition of consciousness can be viewed from two perspectives: the state of being conscious, which equates wakefulness, or the contents of consciousness which includes an awareness of this wakeful state.¹ In this article, the term consciousness has been used to refer to the experience of what it is like to be aware of one's wakeful state.

The conversation about AI's consciousness started with the famous Turing test or 'imitation game' in 1950 while attempting to answer the question of whether a computer can think or not.² Turing proposed a test to assess this: Can a computer, hidden in a room, ever produce outputs that would make the recipient unable to differentiate whether there is a human or computer in that room?² While Turing's main aim was to claim that machines will one day be capable of thinking intelligently, with this test he also argued

in humans.^{6 8} Lastly, the information integration theory of consciousness states that the consciousness of an entity depends on its ability to integrate information so it is not unique to humans and animals.^{6 9}

The other-minds problem stating that we can never know with certainty whether others have consciousness or not, has been an obstacle in the debate of AI's consciousness.¹⁰ Considering this problem, we should be ready for all possibilities including the more liberal views of AI's consciousness coming true in the next few years. Considering the implications of AI's consciousness in the domains where it is applied is pertinent, given the accelerated speed of its adoption.

The use of AI in primary healthcare is an important ground for the conversation of the consciousness of AI systems due to its profound impact on humans' physical and mental health. The Canadian Institute for Health Information defines primary healthcare as services involving 'routine care, care for urgent but minor or common health problems, mental healthcare, maternity and childcare, psychosocial services, liaison with home care, health promotion and disease prevention, nutrition counselling and end-of-life care'.¹¹ Primary healthcare, spanning from infancy to end-of-life stages, is broader and more inclusive than specialty care, making it more comprehensive for AI integration. It addresses diverse health dimensions—mental, physical and social—by focusing on whole person care. Functioning as the front line of healthcare, primary healthcare screens, diagnoses and treats numerous patients, reducing the need for specialty referrals. Especially in rural areas, primary healthcare serves as the main interface between communities and the health system. The extensive patient network of primary healthcare makes it a prime target for the integration of AI.¹²

There is evidence supporting the use of AI for improving many functions within the primary healthcare system including but not limited to prognosis, diagnosis, shared decision-making, resource allocation and policy-making.^{13 14} Despite this evidence, little is known about how an AI system that possesses consciousness, if it were to exist, would support or hinder the current applications of AI in primary healthcare. In this paper, we refer to an AI system that possesses consciousness interchangeably with a conscious AI. In the next few sections, we will provide an outlook of the impacts of conscious AI on each of the domains of AI use in primary healthcare mentioned earlier and will discuss some ethical considerations associated with the use of conscious AI systems.

THE USE OF AI FOR DIAGNOSIS AND RISK STRATIFICATION

A subfield of AI is machine learning, which provides an alternative to traditional statistics for making inferences from data.¹⁵ Deep learning algorithms, being a subcategory of machine learning, process received inputs through hidden interconnected layers superficially similar to neurons and arrive at outputs that help

establish new patterns in data.¹⁵ Some examples of employment of machine learning in primary healthcare are as follows: early detection of diabetic retinopathy,¹⁶ risk prediction for future atrial fibrillation incidents¹⁷ and diagnosis of dermatological disease.¹⁸ Machine learning has been able to exceed human capacities in establishing new associations in large quantities of data.¹⁹ Therefore, AI is already at the verge of revolutionising the way we understand diseases by finding new patterns to classify medical information.

If these AI systems were to possess consciousness, part of our collective medical knowledge would exist beyond the human mind and in the artificially intelligent mind. Medical knowledge extends beyond individual human minds to encompass the collective knowledge of humanity across time. With conscious AI, this boundary expands to include AI minds, comprising all AI systems. Conscious AI may be viewed as part of human cognitive scaffolding or, ambitiously, an extension of the human mind. Consequently, AI transcends being merely a tool or a separate entity from humans; instead, the 'human-machine'²⁰ complex becomes the holder of conceptual ideas about medical pathologies.

However, just like a human healthcare provider, a conscious AI is prone to errors. One factor contributing to making incorrect inferences is the lack of neutrality in AI systems.¹² These systems carry the biases of the humans who design them.¹² Moreover, the data which trains these systems is often biased and reflects inequitable approaches to data collection that have been adopted in healthcare over the years.¹² For example, there is evidence that an AI system designed to estimate the risk of suicide performs better for white patients compared to black patients due to the biases in the data on which it was trained, tested and validated.²¹ If we factor AI's consciousness into this example, by encoding biases into AI systems, we might create conscious entities that are inherently biased and perpetuate the existing disparities of health. To avoid this issue, we need to ensure that data collection processes as well as algorithm deployment and implementation are fair and equitable^{22 23}.

THE USE OF AI FOR SHARED DECISION-MAKING

Primary healthcare physicians often engage in shared decision-making with their patients regarding their care. AI has the potential to improve this process by providing evidence-based information on various alternatives tailored to the needs of each patient.²⁴ However, there are records of reservations from patients towards AI's involvement in the decision-making process.²⁵ Some of the mistrust towards AI stems from the existing challenges in explainability and interpretability of AI processes^{24 25}; in other words, while it is important for patients and healthcare providers to understand the rationale behind AI's recommendations,²⁵ little is known about the exact steps through which AI connects an input to an output.²⁴

By providing an internal awareness of the AI systems' processes, consciousness could enable these systems to critically analyse their outputs. The idea of awareness of internal processes is compatible with the higher order theory of consciousness⁸ and might resemble metacognition. However, metacognition is inherently a conscious process so consciousness can be seen as a prerequisite for metacognition.

If AI systems could critically appraise their conclusions, it might enhance patients' trust in using AI for shared decision-making. Further research is necessary to understand patients' perspectives on AI's role in medical decision-making. Additionally, more studies are needed to gauge how AI's consciousness affects physicians' acceptance and integration of AI into shared decision-making processes.

THE INTERACTION BETWEEN AI IN PRIMARY HEALTHCARE AND HEALTH LITERACY

By involving AI in the care for patients, we should consider the instances where AI would communicate with patients independent of the primary healthcare professional. An example for this comes from the supportive and palliative care setting, an area of medicine mostly run by family physicians in Canada with important implications for community health. The use of machine learning algorithms can increase the accuracy of mortality risk predictions and prompt physicians to have end-of-life planning and goals of care discussions with patients and their families earlier, thus avoiding unnecessary costly interventions.²⁶ However, there is a risk that patients might receive automatic notifications regarding these predictions prior to having a chance to discuss the results with their physician.²⁶ This can negatively impact the patients' well-being and open the door to misinterpretations.²⁶

Patients' health and digital literacy could have an effect on the magnitude of the impact of direct information release from AI to patients. However, currently little is known about the effect of patients' health literacy on their interactions with AI and conscious AI, if any.¹³ An avenue for future research is determining whether there is an association between patients' education level and their attitudes towards predictions made by an AI system and whether these attitudes would change if the AI in question is conscious. Moreover, randomised control trials are needed to look at whether short-term patient education programmes, addressing basic AI principles, would change patients' attitudes towards employment of conscious AI in their care.

PUBLIC HEALTH POLICY-MAKING GUIDED BY AI

AI can be used in primary healthcare for reallocation of resources to better address population's needs. For instance, AI has the power to adjust the number of patients expected to be seen by a single family physician in a certain time frame based on predictions about the

complexities of those patient encounters.²⁷ Moreover, there is evidence that AI can be useful in the redistribution of scarce resources at the hospital or healthcare level during public health emergencies such as the COVID-19 pandemic.²⁸ In addition, the use of AI-enabled point-of-care screening tools in primary healthcare settings has been shown to be cost-effective if a certain threshold of compliance with the tool is met.²⁹ These conclusions indicate that the ability of AI to increase the efficiency of the healthcare system hinges on the physicians', patients' and policy-makers' acceptance of the use of AI. However, it is unclear if AI being conscious would change this acceptance rate.

Some concerns identified from focus groups of patients regarding their apprehensions towards the use of AI include cost-effectiveness, safety, biases in AI systems and the impact of these tools on patient autonomy.²⁵ This shows that the core arguments against the use of AI for policy-making are not about AI's consciousness per se but rather about the sense of control that patients would like to have over their data and how it would affect their faith.

More comprehensive guidelines need to be placed to ensure that collection and use of data by AI respects people's right to privacy and self-determination. The European Commission's regulatory framework on AI proposal serves as a stepping stone in this direction.³⁰ Their risk-based approach to AI systems and proposed additional levels of regulation on higher-risk tasks performed by AI are well justified to ensure the safety of AI use.³⁰ Nevertheless, more clarity is needed regarding the risk classification of current tasks delegated to AI in primary healthcare, and regarding the risk classification of next generation AI systems.

OTHER ETHICAL ISSUES ASSOCIATED WITH THE USE OF AI IN HEALTHCARE

One of the most important ethical issues raised by considering the consciousness of AI is about the responsibilities associated with the status of being conscious.³¹ Bendel discusses the topic of 'machine medical ethics' as applying principles of medical ethics to machines (a term encompassing any form of non-human agents ranging from robots to strong AI systems).³² From this perspective, AI systems are no longer 'objects' of morality but rather 'subjects' of moral principles,²⁰ meaning that they are not just instruments to human agents but that AI systems themselves are bound by principles of ethics for their actions. A great example of applying this framework of ethics in healthcare is to consider where the responsibility of making a diagnostic impression lies when AI systems are employed to arrive at that impression; Is the algorithm itself responsible, the humans who designed it, or the people who used the results of that algorithm?

Some argue that the capacity to form intentions is a prerequisite for placing moral responsibility on AI systems²⁰; AI should be able to understand the consequences of its actions and make a choice with intentions to

be held accountable for its actions.²⁰ However, this brings us to the other-minds problem: We cannot be certain if AI systems comprehend consequences as humans do. Some philosophers might suggest a deontological approach to medical machine ethics. Here, AI would follow programmed values, considering some actions inherently good and others bad. Deviations would breach the machine's ethical principles.³² However, the issue with this perspective is that the AI system becomes dependent on the human-collected data on which it was trained. Therefore, the system's morality aligns with that of its designers, reverting AI to being a tool under human control.

For AI systems to be accountable in medical diagnosis, they require an internal value system to weigh misdiagnosis risks against diagnostic certainty to make recommendations. Presently, this decision-making authority rests with human healthcare providers as the situation might be best described as 'AI augmenting human abilities' in making a medical decision.²⁰ Therefore, the human-machine complex becomes the decision-maker and the subject of ethical principles as a hybrid entity.²⁰ Nevertheless, more interdisciplinary research is needed to investigate the ethics of instances where human or AI exit the hybrid complex and become the sole entity responsible for an action or a medical decision.

CONCLUSION

Current literature supposes that none of the AI systems designed up to now possess consciousness.³ However, it is not far from imagination that we soon might be able to create AI systems that exhibit, or perhaps stimulate, significant human traits.³ AI has shown many promising avenues for improving primary healthcare.^{13,14} However, the success of AI in this field largely depends on healthcare workers', patients' and policy-makers' attitudes towards it and the possibility of AI being conscious has the potential to affect these attitudes and complicate the previously accepted medical ethics principles.

The other-minds problem could constitute a challenge to answering the question of 'whether AI has consciousness' in a concrete way. Nevertheless, the key to solving the ethical dilemmas raised by this question might be to shift our perspective to the human-machine complex as the subject of ethical principles. This approach would require that the work of AI developers, AI users, as well as the overall output of the human-machine entity to be ethical.

A strength of this commentary is that it targets the novel topic of the consciousness of AI in the context of primary healthcare. To date, there is no work reflective of this unique perspective in the literature and this paper serves as a steppingstone for further integration of the theories of consciousness of AI with the use of AI in healthcare. A limitation of this commentary, however, is that due to the scarcity of the established evidence on the topic of consciousness of AI and its application to primary healthcare, most of the content of this paper reflects what authors were able to find in the existing literature.

As a direction for future work, to better understand the human-machine complex, more research must be done on its intrinsic relationships. For example, we need to assess how considering AI as a conscious being would change healthcare providers' and healthcare receivers' attitudes towards its use in practice. With more interdisciplinary research on the implications of conscious AI in its application in healthcare settings, we might be able to create more guidelines for the human-machine healer to act ethically on micro to macro levels. Lastly, further research is needed to extend our understanding of human consciousness which could lead to not only improving AI systems to better mirror or complement human cognition, but also could eventually lead to treating different diseases including neurological disorders.

Acknowledgements The authors would like to acknowledge that Professor Jocelyn Maclure provided comments on the content of the paper during the revision process, and these comments were incorporated into the final version of the manuscript. SAR is Canada Research Chair (Tier II) in Advanced Digital Primary Health Care, received salary support from a Research Scholar Junior 1 Career Development Award from the Fonds de Recherche du Québec-Santé (FRQS) during a portion of this study, and her research program is supported by the Natural Sciences Research Council (NSERC) Discovery (grant 2020-05246).

Contributors DR: investigation, data curation, writing-original draft, writing-revisions and edits. SAR: conceptualisation, methodology, supervision, project administration, writing-revision and edits.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Samira Abbasgholizadeh Rahimi <http://orcid.org/0000-0003-3781-1360>

REFERENCES

- 1 Zhao T, Zhu Y, Tang H, et al. Consciousness: new concepts and neural networks. *Front Cell Neurosci* 2019;13:302.
- 2 Turing AM. Computing machinery and intelligence. *Mind* 1950;59:433–60.
- 3 Schwitzgebel E. AI systems must not confuse users about their sentience or moral status. *Patterns (N Y)* 2023;4:100818.
- 4 Searle JR. Minds, brains, and programs. *Behavioral and Brain Sciences* 1980;3:417–24.
- 5 Graziano MSA, Webb TW. A mechanistic theory of consciousness. *Int J Mach Conscious* 2014;06:163–76.
- 6 Friedman G, Turk KW, Budson AE. The current of consciousness: neural correlates and clinical aspects. *Curr Neurol Neurosci Rep* 2023;23:345–52.
- 7 Mashour GA, Roelfsema P, Changeux J-P, et al. Conscious processing and the global neuronal workspace hypothesis. *Neuron* 2020;105:776–98.
- 8 Brown R, Lau H, LeDoux JE. Understanding the higher-order approach to consciousness. *Trends Cogn Sci* 2019;23:754–68.
- 9 Tononi G. An information integration theory of consciousness. *BMC Neurosci* 2004;5:42.

- 10 Harnad S. Other bodies, other minds: a machine incarnation of an old philosophical problem. *Minds Mach* 1991;1:43–54.
- 11 Canadian Institute for Health Information. Primary care - Canadian institute for health information. Available: <https://www.cihi.ca/en/topics/primary-care> [Accessed 20 Dec 2023].
- 12 Lin S. A clinician's guide to artificial intelligence (AI): why and how primary care should lead the health care AI revolution. *J Am Board Fam Med* 2022;35:175–84.
- 13 Abbasgholizadeh Rahimi S, Légaré F, Sharma G, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. *J Med Internet Res* 2021;23:e29839.
- 14 Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. *J Gen Intern Med* 2019;34:1626–30.
- 15 Tolsgaard MG, Pusic MV, Sebok-Syer SS, et al. The fundamentals of artificial intelligence in medical education research: AMEE guide No.156. *Med Teach* 2023;45:565–73.
- 16 Wewetzer L, Held LA, Steinhäuser J. Diagnostic performance of deep-learning-based screening methods for diabetic retinopathy in primary care—A meta-analysis. *PLoS One* 2021;16:e0255034.
- 17 Khurshid S, Friedman S, Reeder C, et al. Electrocardiogram-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation* 2022;145:122–33.
- 18 Jeong HK, Park C, Henao R, et al. Deep learning in dermatology: a systematic review of current approaches, outcomes, and limitations. *JID Innov* 2023;3:100150.
- 19 Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med* 2022;28:31–8.
- 20 De Togni G, Erikainen S, Chan S, et al. What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care. *Soc Sci Med* 2021;277:113874.
- 21 Coley RY, Johnson E, Simon GE, et al. Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. *JAMA Psychiatry* 2021;78:726–34.
- 22 Nyariro M, Emami E, Caidor P, et al. Integrating equity, diversity and inclusion throughout the lifecycle of AI within healthcare: a scoping review protocol. *BMJ Open* 2023;13:e072069.
- 23 Nyariro M, Emami E, Caidor P, et al. Integrating equity, diversity and inclusion throughout the Lifecycle of AI within Healthcare: a Scoping review protocol. *BMJ Open* 2023;13:e072069. Available: <https://doi.org/10.1136/bmjopen-2023-072069>
- 24 Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, et al. Application of artificial intelligence in shared decision making: scoping review. *JMIR Med Inform* 2022;10:e36199.
- 25 Richardson JP, Smith C, Curtis S, et al. Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med* 2021;4:140.
- 26 Parikh RB, Manz CR, Nelson MN, et al. Clinician perspectives on machine learning prognostic algorithms in the routine care of patients with cancer: a qualitative study. *Support Care Cancer* 2022;30:4363–72.
- 27 Rajkomar A, Yim JW, Grumbach K, et al. Weighting primary care patient panel size: a novel electronic health record-derived measure using machine learning. *JMIR Med Inform* 2016;4:e29.
- 28 Wu H, Lu X, Wang H. The application of artificial intelligence in health care resource allocation before and during the COVID-19 pandemic: scoping review. *JMIR AI* 2023;2:e38397.
- 29 Wolf RM, Channa R, Abramoff MD, et al. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol* 2020;138:1063–9.
- 30 European Commission. Regulatory framework for artificial intelligence - digital strategy. Available: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [Accessed 20 Dec 2023].
- 31 Masters K. Ethical use of artificial intelligence in health professions education: AMEE guide No.158. *Med Teach* 2023;45:574–84.
- 32 Bendel O. Surgical, therapeutic, nursing and sex robots machine and information ethics. In: Van Rysewyk SP, Pontier M, eds. *Machine Medical Ethics*. Cham: Springer, 2015: 17–32.