International Journal of Digital Health
Original Article Open Access

We Must Not Hide Behind the 'Black Box' Problem as a Reason to Reject or Avoid Adopting A.I. Technologies": A Qualitative Study with Canadian Healthcare Professionals Regarding AI use in Healthcare

Sarah Bouhouita-Guermech[1,2*] and Hazar Haidar[3]

[1]*School of Public Health, University of Montreal, 7101 Park Ave, Montreal, Quebec, H3N 1X9, Canada*
[2]*Centre of Genomics and Policy (CGP), McGill University, 740, avenue Dr. Penfield, suite 5200, Montreal, Quebec, H3A 0G1, Canada*
[3]*Ethics programs, Department of Letters and Humanities, University of Quebec at Rimouski, 300, allée des Ursulines, C.P. 3300, succ. A, 300 All. des Ursulines, Rimouski, Quebec, G5L 3A1, Canada*

Abstract

Background: Artificial intelligence (AI) has had a profound impact across various sectors worldwide, including education, marketing, and healthcare, and its influence continues to grow. In healthcare, AI is considered revolutionary for its potential to significantly improve diagnosis, treatment, and care planning. Canada, a leader in AI adoption, is at the forefront of initiating AI implementation strategies. However, despite its benefits, AI raises several ethical challenges, regarding data privacy, explainability, and bias, which can hinder its adoption. In healthcare, where clinicians and patients are the primary users, understanding their perspectives is essential for fostering AI acceptance and ensuring its ethical integration.

Methods: We interviewed 15 healthcare professionals across Canada to gain insight into their views on AI in healthcare, their interactions with AI (if any), and their expectations for its future. Additionally, we invited them to reflect on possible challenges regarding AI in healthcare.

Results: Clinicians had varying perceptions towards AI use in healthcare, ranging from a positive and promising attitude to a more skeptical one. Many acknoweldged its potential benefits, such as reducing administrative tasks and improving diagnosis. However, all participants expressed concerns about AI use in healthcare, particularly regarding privacy issues and the validation of data and AI systems.

Conclusions: While AI promises significant improvements in patient care, our findings reveal that this potential is accompanied by notable concerns. Healthcare professionals would greatly benefit from AI education tailored to their medical practice, along with clear guidelines that enhance their confidence and capability to use AI in healthcare.

Publication History:

Received: June 18, 2025
Accepted: July 20, 2025
Published: July 22, 2025

Keywords:

Artificial intelligence, Healthcare professionals, Ethics, Qualitative, Interviews, Canada

Introduction

Artificial intelligence (AI) has made, and continues to exert, a profound impact across various sectors globally, including education, marketing, and health [1]. Broadly defined, AI is the field of computer systems designed to replicate human activities, ranging from performing programmed tasks to conducting predictive analysis to aiding in decision-making [2]. For instance, in marketing, AI models predict customers purchasing tendencies and generate insights for optimizing a company's website [3]. In agriculture, AI technologies can boost crop yields by providing more accurate predictions for optimal crop selection, planting times, and harvesting locations [4].

In healthcare, AI's potential is seen as revolutionary, with the ability to significantly enhance the healthcare landscape by improving patients' diagnosis, treatment, and care planning [5]. AI is expected to proliferate the accuracy of AI models and their application in medical diagnostics through prediction analysis and assistance for healthcare professionals (HPs) in their medical decision-making for an optimal patient care [6]. For example, AI has been used to identify protein complexes leading to potential therapeutic targets [7] and to assess an individual's risk of developing specific diseases [8]. While the integration of new technologies into healthcare systems can be slow due to the field's risk-averse nature, the advent of practical and portable devices, such as smartphones, and the growing availability of digital records for AI training have improved AI implementation [9–11].

Many countries are making progress in implementing AI across diverse sectors. Canada, a leading example, has been at the forefront of initiating AI implementation strategies, pioneering the first national AI strategy in 2017 and remaining actively engaged in this field ever since [12]. As part of these initiatives, Canada has established key centers to promote and advance AI research and implementation such as the Vector Institute in Toronto, Ontario; Mila - Québec AI Institute in Montréal, Québec; and Alberta Machine Intelligence Institute (Amii) in Edmonton, Alberta [12].

There is a notable increase in AI investments, such as in Ontario, which promise to enhance healthcare services, among other sectors [13]. Further, Canada is home to several companies that have developed AI-driven medical devices, including GE Healthcare's Critical Care Suite, which aids radiologists in performing medical imaging more efficiently through an AI-integrated mobile device [5, 14]. Another example is Swift Medical, a Canadian healthcare

*Corresponding Author: Sarah Bouhouita-Guermech, School of Public Health, University of Montreal, 7101 Park Ave, Montreal, Quebec, H3N 1X9, Canada

Citation: Bouhouita-Guermech S, Haidar H (2025) We must not Hide Behind the 'Black Box' Problem as a Reason to Reject or Avoid Adopting A.I. Technologies": A Qualitative Study with Canadian Healthcare Professionals Regarding AI use in Healthcare . Int J Digt Hlthc 2: 112. doi: https://doi.org/10.15344/ijdh/2025/112

technology firm that leverages AI to assist clinicians in analyzing and managing wounds effectively [13].

Despite its benefits, AI raises several ethical concerns, including those related to data privacy, as the reliance of AI systems on large datasets complicates data protection, leading to potential invasion of privacy and data misuse [15]. Another widely debated ethical concern is AI explainability commonly known as the black-box problem, which refers to the difficulty in understanding the rationale behind the AI system's generated outcomes [16]. Additionally, AI bias often results from unrepresentative data in current datasets, leading AI models to produce outcomes that might not be appropriate for certain populations and potentially resulting in issues such as misdiagnosis [17]. These concerns can create hesitancy among users and stakeholders, thereby hindering AI adoption. In healthcare, where clinicians and patients are the primary users, understanding their perspectives about AI is crucial for fostering its acceptance [18].

This paper explores Canadian HPs' views regarding the current and future applications of AI in their respective fields. We interviewed 15 HPs to gain insight into their perspectives regarding the use of AI in healthcare, interactions with AI, if any, and their expectations for its future. Additionally, we invited them to reflect on the challenges that AI may pose concerning its integration and use within the healthcare system. Its noteworthy that throughout the interviews, AI was used as an umbrella term to refer to any tool or system that uses large datasets to assist in medical practices, such as diagnosis, treatments, and triage, among others.

Methodology

Study design

We used a qualitative description (QD) methodology, adopting a semi-structured interview approach with HPs across Canada. QD is particularly valuable when exploring nuanced topics such as the use of AI in healthcare. It allows for a comprehensive understanding of participants' perspectives and experiences by capturing HPs' insights into the issues and concerns raised by AI, as well as the challenges they might encounter in their practice [19, 20].

Sampling and recruitment

Participants were included in this study if they were HPs with current or past practice experience in Canada, are using an AI tool in their practice or have an interest in AI. For the recruitment phase, we circulated informational flyers via email and by posting in various institutions such as universities and hospitals. HPs' emails were curated from publicly available information (i.e., websites) and interested participants were invited to contact us by email or phone. Recruitment continued until data saturation was attained and interviews were conducted from July 2022 to September 2023. Participants who expressed interest by responding to our email invitations or follow-ups were asked to provide their availability to schedule the interviews along with their signed consent forms. In total, 15 HPs across Canada were interviewed for this study. Figure 1 shows a detailed participant demographic information. The duration of the interviews ranged from 30 minutes to 1.5 hours and were recorded through Zoom.

Data collection

Semi-structured interviews were selected as the data collection method, allowing to guide the interview while engaging in open-ended discussions about the topic. This approach facilitated elaborate discussions with participants who were able to express their views and concerns, contributing to a more comprehensive understanding of the issues they raised [21]. Two independent researchers (SBG and HH) conducted semi-structured interviews with the recruited participants across Canada. SBG and HH developed the interview guide following an overview of the literature related to the topic. The interview guide explored diverse themes including the integrity of AI-generated data, accountability and the patient-clinician relationship, free and informed consent, trust and transparency, and education. The interview guide is provided in Appendice 1 (supplementary file). Interviews were audiotaped, transcribed verbatim, and anonymized.

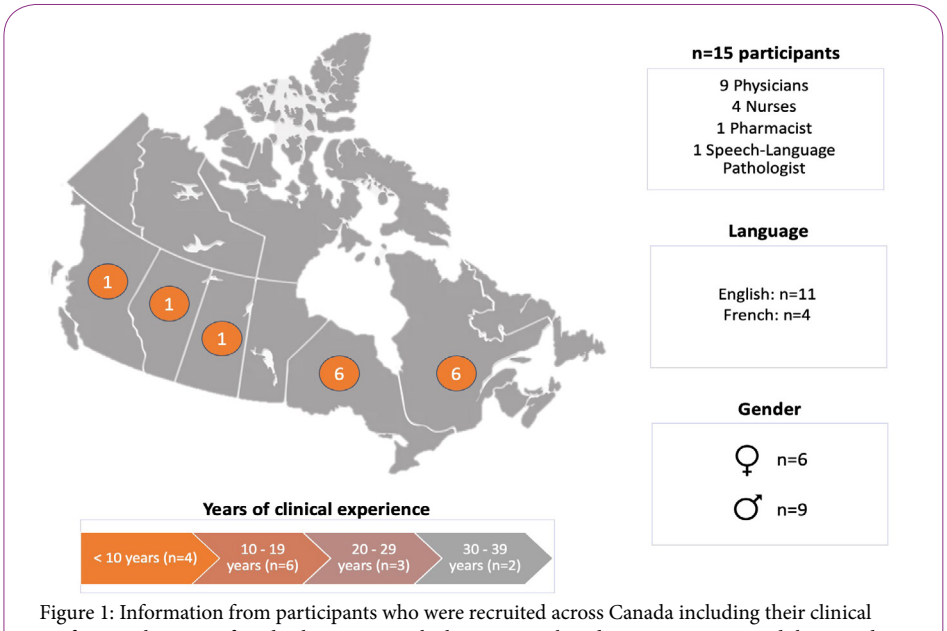

Figure 1: Information from participants who were recruited across Canada including their clinical profession, the years of medical experience, the language spoken during interviews, and their gender.

Data analysis

We used a reflexive thematic analysis for our data analysis to identify emerging themes and perceptions [22]. SBG and HH analyzed the total transcripts using NVivo 12 software and extracted prominent themes to develop a codebook [23]. The codebook was established through a process involving the review and coding of transcripts and completed after conducting a series of discussions to reach an agreement on the identified themes and their definitions. Subsequently, SBG and HH analyzed each transcript using the finalized codebook as a guide.

Results

Key findings

Participants came from diverse professional backgrounds as well as varying provinces, allowing for different contextual intakes on the topic. Around half of the participants held extensive knowledge and first-hand experience with AI in healthcare allowing them to share their reflections with specific scenarios, while the other half exhibited a more general interest in the matter and tended to answer hypothetically. Our results section will present varying perceptions HPs have towards AI use in healthcare, ranging from an optimistic attitude to a more skeptical one. Additionally, we report on the main benefits and concerns that participants have expressed regarding AI use and integration in healthcare. Some of our participants shared examples of AI systems that they have used or are currently using during their practice, which are listed in Table 1.

HPs' perceptions and attitudes towards AI use in healthcare

Throughout the interviews, participants conveyed diverse perspectives. While the majority recognized the significant benefits AI could offer to healthcare in general, all of them expressed caution regarding its potential risks. Although some HPs were more skeptical than others, participants generally exhibited a sense of ambivalence toward the use of AI in healthcare. Further, despite maintaining curiosity about AI, there was widespread reluctance to fully trust these new and unfamiliar technologies. This following section illustrates the positive, negative, and precautionary attitudes that emerged from the interviews with HPs.

Purpose of AI Model	Description
Rare Disease Diagnosis	The AI model analyzes patient information (i.e., lab tests, consultation notes, medications) to signal individuals who are at high risk of rare disease diagnosis.
Transcription	The AI model analyzes the patient-clinician consultation to generate a medical note for the physician.
Chatbot (i.e., Specific Clinical Chatbots, ChatGPT)	The chatbot analyzes the medical question and generates a response according to the available data.
Workflow Optimization Algorithm	This algorithm assists clinicians with their administrative tasks.
Decision-Making Tool	This tool helps clinicians in choosing the right treatments for their patients.
Predictive Analytic Tools	Some of these tools can be useful to emit an early signal when a patient is at risk of deteriorating or dying and is in need of intensive care. Other tools are used to predict whether a patient is ready to be discharged from the hospital.

Table 1: Examples of AI use in medical practice as reported by participants

Many of our study participants highlighted AI's rapid growth and the significant impact it has already made in healthcare, noting its swift integration and early contributions for its users. As one HP mentioned, "it's already helping with pattern recognition, vision recognition, and in terms of diagnostics' improvement" (HP11). These highly anticipated AI models can amass and process large amounts of data which, many participants pointed out, could allow more accurate results predictions (i.e., diagnostics, medical outcomes, waiting times, influx of patients into ERs, drug intake), thereby improving healthcare systems by yielding greater diagnostic precision and enhanced health management.

> *" [AI] made research way easier. You can collect data and analyse data way easier, which would advance the medical field, definitely. It makes your practice way safer for patients, especially in, like, calculating medication doses, dosages, especially for pediatric patients, because it depends on the weight. So, you would have it automatically. Allergies, so you don't order any medications for which the patient has an allergy. And so on and so forth. So, I think it's […] definitely safer, it's definitely better for patient care. And this is only in the emergency. But, even, like in surgical care, in everything, it's way better and we should all embrace it. But know how to embrace it"* (HP3).

The majority of the participants shared their enthusiasm for AI's future advances and integration in their practice. Some HPs discussed the way AI's performance will continue to improve and that it is predominantly perceived as an ongoing process, portraying that AI's potential is still evolving and becoming more sophisticated.

> *"We're just beginning. And, as these AI engines get smarter, and we start to trust them more, we're going to be able to help have the AI engines help us to produce better data"* (HP14).

> "*It's going to evolve, regardless. And, it's going to evolve just like the Internet did. And, have lots of really great things, and lots of seedy things, in the future*" (HP4).

The progress of AI is expected to continue, compelling clinicians to learn to use it. This may impact HPs' medical practice, depending on how readily they would adopt the technology. As a participant emphasized that "a doctor […] who won't adopt artificial intelligence, is going to fall behind" (HP10).

Some participants felt more wary about AI's dissemination and claimed that "we are still in the stone age" (HP6) for this technology to be adequately used and appreciated. Some clinicians do not perceive AI as having a significant impact yet, while others are addressing the considerable delays in implementing AI in the healthcare system and the slow pace at which new technologies are generally integrated, particularly in Canada. As one participant stated:

> *"We've already really, functionally missed the bus in being part of the discussion on what can be, what should be, what could be developed. Now, we haven't missed it, completely. But we're missing a lot of it, right now. Because there's people, and there's companies out there*

that're just doing it for us, while we're still trying to install, in scale, an electronic health record. Like, we're in 2022, and we're still trying to install electronic health records, in Canada. And, we have companies out there, like Amazon, who just go out and buy, in the Walmarts, various health clinics, in scale […]. [We're] 20 years behind. There's, even, articles on nursing informatics, and health informatics that have been published recently, that show you the uptake across Canada. And it's not good. In 2021, 2022" (HP4).

AI implementation and use are not linear processes; many factors should be considered before adopting a new AI method or tool in clinical care. Several participants explained that the adoption of a new practice or tool occurs in stages. One HP provided an insightful observation on how clinicians tend to approach a new AI tool, stating the following:

*"First, distrust; second, curiosity: "So, show me;" third, "Show me from the point of view of the quality, not from the point of view of punishment or regulation or accountability;" […] fourth, "Give me some tools to improve;" fifth, you know, improve my work-life balance and my experience of being a clinician" (*HP14).

One HP expressed being "[n]ot confident, [b]ut looking forward" (HP15) to the new technology, reflecting a nuanced perspective shared by many participants. This perception reveals both optimism and curiosity about the technology's potential, alongside hesitation, as HPs feel unequipped and unprepared to fully embrace it. Indeed, most participants agree that implementing AI technologies must be accompanied by a rigorous validation process, which is essential for fostering trust in AI:

"I think there'll always be physicians like me, who're early adopters, who will feel interested and excited to take a little bit of a risk and experiment with tools that haven't been completely, fully regulated and validated by, you know, institutional organizations like Health Canada. But I think what is required to get, you know, universal buy-in and adoption, and hopefully, transformation in the system will be those robust regulations in place, doing what they're supposed to do. And, hopefully, that will lead to a proliferation of AI-based tools that can really help healthcare" (HP2).

Most clinicians do not think that all new AI models should be adopted as the standard of care at this stage, despite their promising outcomes. AI technology is either still in development or not easily adopted by all groups of clinicians, raising questions about its acceptability in healthcare. One HP highlighted this concern by noting that:

*"Accuracy of some of these models is the big issue, in the short term. I'm sure they'll get our trust later on. […] Because, honestly, it's still faster to access things that you know, rather than just going to some of these AI tools. […] There's a lot of physicians […] who when they learn with ChatGPT and they get shown all the use cases, they actually still don't want to use it. So, I don't think they should be forced to use this technology even though it's helpful for however many people there are" (*HP12).

AI benefits according to healthcare professionals

Our study participants have acknowledged and reported the current and potential benefits of using AI in healthcare, including a) alleviating administrative tasks, b) reducing waiting times, and c) improving diagnostic outcomes.

a. Alleviating administrative tasks

Many participants agreed that integrating AI into medical practice could potentially alleviate the burden of heavy administrative tasks HPs regularly face, namely "clinical documentation" (HP11), extensive "data analysis" (HP10), and score calculations (HP3, HP8). As one clinician pointed out, "AI has the potential to offload a lot of the administrative effort, and therefore – of clinicians – and therefore, free them up to do what they're actually trained to do" (2). Further, HPs were hopeful that reducing administrative work could improve their relationships with patients by allowing more time for patient-HP interaction.

"[It] could improve, by reducing the need for, sort of, manual documentation, and all of the hours that are spent on that, it could actually improve the amount of time left for human interactions. And improve the quality of those interactions" (HP11).

"People are sometimes afraid that AI will make healthcare or medicine less humane. But, I think, on the contrary, it will allow us to focus on what's most human; and that's spending time with our patient and answering their questions, again, without worrying about the lack of time" (HP10).

b. Reducing waiting time

Many HPs mentioned the potential of AI to improve healthcare delivery by decreasing waiting times for patients. According to one participant, AI can reduce waiting times by "better direct[ing] individuals to the respective healthcare professionals" (HP7) based on their medical condition. Further, another HP mentioned that AI could streamline "the time it takes to diagnose an illness that isn't obvious right off the bat" (HP1), such as ambiguous or rare diseases, and "can improve the time and length of procedures, when it's time to forward requests, to send results" (HP9).

c. Diagnostic improvements

AI can help improve overall healthcare delivery by enhancing diagnostic accuracy and therefore reducing waiting times. Many HPs have noted that advancements in AI have already contributed to improvements across various areas of healthcare.

"[In radiology], the algorithms that are currently being, well, evaluated are actually proving to be more accurate and therefore less medical errors are going to occur from them than just the clinicians on their own" (HP2).

"[It's] already helping with pattern recognition, vision recognition, and in terms of diagnostics improvement" (HP11).

With AI's ability to provide more accurate diagnoses, diseases can be detected at earlier stages, enabling timely treatments and improving patients' chances of a healthier life. One participants highlighted that AI is valuable in effectively identifying high-risk patients, which can accelerate interventions: "*So, rather than having to scramble at that point at the three-hour mark, once the decline happened, all those teams were already aware and were ready to mobilize very quickly and that patient did quite well because of that very quick mobilization*" (HP2). Further, according to HPs, AI's effectiveness in processing large datasets could not only enhance

diagnostic accuracy but also bring us closer to achieving "personalized medicine" (11). This would enable physicians to "look at the information, the data that a patient has and monitor changing parameters, to fine-tune their management". (1)

Participants also mentioned additional benefits of AI in healthcare, such as its effectiveness in "knowledge synthesis" (HP15). Many HPs deal with vast amounts of constantly growing new information and need to stay updated. Hence, AI's contribution to knowledge synthesis could be highly beneficial for HPs, who are often expected "to incorporate all the latest information and research and knowledge that's required for good clinical care" (HP14).

Others stated that, through AI implementation, it might be possible to establish more standardized medicine by potentially closing the gap between different practices adopted by clinicians, which could lead to more equitable healthcare: "*[If] well used, [AI] could also help to standardize decision-making a little, so that there aren't too many big differences - sometimes - between healthcare professionals" (7).*

Ethical challenges in the management of health data

Most concerns reported by participants were related to data use. While several general issues with using AI were raised, data-related concerns -spanning both inputs and outputs throughout AI lifecycle- were particularly prominent in the interviews. The following section summarizes the ethical challenges associated with data, including data misuse, data bias, privacy issues, the AI black-box problem, and validation.

Data misuse

Many participants not only highlighted the importance of data, describing it as "the newest form of currency" (HP1), but also emphasized the need for responsible data usage. With growing concerns about selling data to commercial companies or other entities that could exploit individuals' information, one participant stated the following:

*"Making sure people know when to use, when they can use patient information and when not. Making that very clear in the regulations. Making sure companies aren't holding a lot of patient information. And as more of this data gets accumulated with all the EMRs and what not, it's kind of like whether or not a clinic can use the data they have. And, I mean, can they monetize that data by selling it to a company, to train their models?" (*HP12*).*

Health data is continually being collected to train more accurate AI systems. However, as data collection and AI development advance, it is becoming increasingly difficult for the public and experts to keep track of this data and ensure its responsible use. Many HPs are concerned about the potential misuse of this data, fearing it could be used to manipulate individuals by interfering with and/or influencing their daily decision-making.

"[When] you have technology that can sway behaviour based on people's information that they've either passively been collecting, or personally entered into with a system, to then make decisions on their behalf. […] That is what scares me more about anything else, especially in terms of the patient. And, related to healthcare. We've seen it happen in the social spectrum, in terms of people's ideas and beliefs around, you know masks, and vaccines and everything else under the sun. Tide Pods and… ugh. But, you know, like, it's… when we start taking people's personal health information and doing it to them, what's to stop companies from monetizing behavioural change in their favour?" (HP4).

Data bias

One significant concern with AI tools is the generation of biased outputs. When discussing AI bias, many participants pointed that such bias stems from the use of unrepresentative data to train AI systems. This biased data can lead to unfair and detrimental outcomes, particularly for minorities whose data is inadequately represented. An example of a negative outcome that could arise from data bias is the unfair prioritization of one group of individuals over another when allocating resources (HP12).

"You know, one thing that comes up often is though, like, the data that some algorithms are trained on are not well-balanced to represent all, you know, groups particularly marginalized groups, under-represented groups, and that can lead to algorithms making predictions that are not appropriate for those groups. Because the data was just not there, right?" (HP2).

One participant believes that this issue could be mitigated through AI's learning functions: "*The good thing with AI, like true AI, not just algorithms, is that it evolves, it evolves all the time. Sure, there's going to be bias, but I think the bias is going to be less with good AI, especially in the future*" (HP3). However, most of our interviewees think that AI will continue to perpetuate bias as long as it is fed with unrepresentative data. While most participants agree that HPs should be able to intercept and correct biased results, they also note that these expectations are not always met. They emphasize the difficulty in ensuring that all biases are identified, particularly given the various types of data bias, which can make it overwhelming for any single individual to manage.

"It's very difficult for doctors, in practice, to recognize these biases or errors, because you have to remember that these predictions are made on a case-by-case basis. So, one patient at a time. And so, our mental capacities don't allow us to have a global picture of the efficiency of these systems, to be able to detect systemic biases" (HP10).

While HPs should be trained to recognize biases and incorporate patients' specific needs and conditions into their decision-making processes, other stakeholders also need to be involved in reducing risks of AI bias, including hospitals (HP10) and governments (HP2). Many participants urged stakeholders to ensure that more data representing the local population is gathered and used to train AI systems. One participant outlined three conditions to help reduce AI bias:

"We need to ensure that these models are trained on databases that contain populations representative of our population and our clinicalreality. Second point, yes. Make sure we develop models based on databases of diversified populations. And, thirdly, by using these models and ensuring that these biases are absent, we can ensure that the outputs of these results do not disadvantage certain ethnic groups" (HP10).

Challenges in safeguarding patient data

Many discussions focused on data sharing and access. The issue of consent was a primary concern, with participants emphasizing the

importance of not only obtaining consent from individuals before collecting their data but also redefining what consent means in the context of AI.

"The aggregation of different types of data into one platform for training AI models [is the critical foundation which] patients have to understand that that is likely going to be occurring, and the system is actually wrestling about how patient consent for how their data is used for that purpose is going to take place is like, that's – it's not exactly clear, you know? How much is it direct consent that's required or is it implied consent, given that it's, you know, for the net positive of population health, right?" (HP2).

Further, many participants highlighted the ambiguity surrounding data management, including where the data should be stored and who should have access to it. The principle of consent is emphasized in respect to data ownership, asserting that data provided by a user belongs to them.

*"And the data is theirs by – in terms of what they – it was built on their backs, with their labour. It's owned by the patient but stewarded by physicians […]. I don't think we should have access to it without consent. So, we have a lot to think about with data normalization, data use and consent. I think we can handle consent, easily, by going directly to the patients, and we can do that by showing them the benefit of that – of their data being used for good" (*HP14*).*

However, one participant cautioned that even if data becomes accessible to patients, it does not necessarily ensure their understanding of the information:

"What I see is that people have access to it, but they don't have the knowledge to analyze the results. And I don't think technology is going to provide that knowledge." (HP6).

Clinicians' attitudes towards the AI black box problem

The AI black box issue was frequently discussed during the interviews. The level of understanding and the perceived impact of this notion varied among clinicians, with diverse opinions on its significance for medical decision-making.

Most HPs identified the AI black box as problematic, with some emphasizing the importance of fully understanding the process behind a tool that is being used before considering the generated results. According to these HPs, ensuring transparency is crucial for identifying and addressing potential errors made by the AI system.

"I mean, it's critically important, right? We don't surrender our decision-making in healthcare, easily. With good reason. You know, in nursing, you know, the level of expertise that you develop as a practitioner, over many years, is really one of the critical foundations of our healthcare system. So, the 'black box,' you know, there's probably no other good analogy, where we would say, "Well, we don't really know how – what's going on in there. But we're just going to go…" Like, that's not – healthcare is very risk-averse. I can think of no other example where I would be like, even in the lab, we are, down to a cellular level, we understand what is happening, there. Understanding is the foundation of healthcare delivery" (HP5).

Surprisingly, some did not view the AI black box issue as significant as it is often portrayed by AI researchers. Reasons for this perspective include the unfeasibility of continuously providing detailed explanations for every aspect of a medical diagnosis or treatment and the belief that, as HPs, they are sufficiently equipped by their training and experience for clinical decision-making. Moreover, there is a strong expectation that HPs will apply their clinical judgement with any tool they use, as they ultimately should make decisions based on their own reasoning.

"If you have the knowledge, have the domain expertise, you could actually figure out what actually is going on, on a technological standpoint. […] I don't think medical people care as much about the 'black box' concept as, maybe, some researchers might think, because there's a lot of things we use and in practice that – you shouldn't be blindly following things, anyways. You should be using your clinical judgement a lot of places. So, at the end of the day it's kind of still on you, what decision you make, regardless of whatever tool you use, it's using a 'black box' or not" (HP12).

"We must not hide behind the 'black box' problem as a reason to reject or avoid adopting A.I. technologies" (HP10).

While most participants acknowledge the existence of the black box problem, some have suggested ways to mitigate this issue by encouraging AI users and developers to maintain a human-in-the-loop model to filter and interpret results, based on each individual's capacity to understanding AI. They also call for further research into AI explainability to address black box dilemma.

"AI – one of the things that it should be doing more and more is actually providing human-level justifications. But it is, based on these black box – very complicated neural network, you know, probability evaluations – that it can't – it's not in its nature to provide very clear reasons as to why it's predicting the way it's predicting, right? So, I think that is an area of research, to try and address that. But, in the meantime, there has to be this interplay between human expertise and the AI's expertise. And, coming to the patient with what is occurring and how we can trust a system that we can't fully scrutinize" (HP2).

Therefore, participants highlighted the importance of not relying solely on AI, but rather use it as a decision-aid tool. Many emphasized the need for critical reflexive thinking when employing these tools, noting that excessive dependence on them without questioning their outputs can be problematic.

"*It just needs to be right enough times for you – to convince you to use it. And then, once it's changed your behaviour, then you're going to continue to use it and not question it. And that's where the danger is – of the 'black box' comes in*" (HP4).

Validity of AI systems' outputs

Many clinicians consider AI tools approved by Health Canada to be safe for use. However, even with approval, these tools are still prone to making errors, prompting HPs to carefully verify the results. Validated AI-generated results involves numerous factors, and many HPs have repeatedly asserted that the validity of AI systems and tools is primarily affected by the quality of the data, particularly if the data contains erroneous information.

" An AI tool needs to be very well-validated, because there's this element of, you know, being very powerful, right? And so, one might think that, you know, if it's such a powerful tool, it's able to synthesize so much data – naturally it has to be valid. But that's not the case. It might not be, necessarily, the case, right? So, that's, I think, one of the big challenges of AI" (HP15).

Use of more resources when unsure of results

While many HPs are committed to evidence-based practices, some believe that solely relying on evidence-based procedures is insufficient. They advocate for an approach that regularly revisits the tool, seeks second opinions, and remains patient-centered.

"I might take aspects of it, I might add a second test, I might consult a colleague… essentially, you know, one of the things that I think we really need to lean into is a better understanding of what we mean by evidence-informed practice, as opposed to just the blind acceptance of evidence-based practice where you, you know, apply the same check-list every time" (HP5).

Importance of user feedback of AI tool

While the validation of medical devices, including AI tools, is a shared responsibility among diverse stakeholders such as developers and governments, all participants agree that HPs have a a crucial role in validating AI-generated result, given their primary responsibility for patient care. Some participants suggested incorporating user feedback into the validation process, as they are well-positioned to identify flaws that occur in practice. This approach could be applied to both approved and non-approved AI medical tools, as these tools cannot be considered completely accurate or reliable enough to be established as the standard of care.

"If there is an error at the front line, of the AI output that's, like, very clear, I'm very in favour in a direct feedback that the clinician can make to the collaborative, of the vendor who made the tool, and that government body, where the centralized AI is occurring. Because, how to unpack how the error occurred is actually very complicated computer science, right? Because either the data that the model is trained on is not, has not been fully optimized. Or, the actual model development could be improved – the parameters of the model. Yes, they got through the Health Canada process, but they still can, …, it's not 100%. There's still always room for improvement" (HP2).

Most clinicians agree that more needs to be done to ensure AI is properly validated. Efforts to use AI responsibly should not rest solely on individuals; governments must also take action to regulate the use of these emerging tools.

"It always comes back to regulation. So, validate our systems before using them despite the hype linked to artificial intelligence. So, never forget the - this first point" (HP10).

Discussion

In this study, we report the perceptions of Canadian HPs about AI integration into healthcare and the potential concerns it may raise. Some participants had prior experience with AI tools (Table 1), whereas others had not yet encountered these systems in their current practice. Participants in our study generally recognized AI's potential to revolutionize healthcare, particularly through its ability to reduce clinicians' administrative workload, allowing them more time to focus on patient care, and to decrease waiting times by enabling more efficient triage and diagnosis. However, they unanimously expressed concerns about its broader implications. Some were particularly cautious, warning of the risk of AI dependency, where clinicians could become overly reliant on the technology, and questioning whether it has been overly hyped or inadequately integrated.

Fostering trust in AI for better adoption

Clinicians' attitudes toward AI vary widely and seem to be influenced by different factors. For instance, our interviews revealed that those with greater experience or exposure to AI tools were more inclined to recognize AI's potential and advocate for bolstering regulation, rather than dismissing its use altogether. Clinicians who understand AI's potential are then more likely to trust its outputs, as highlighted in Shamszare and Choudhury's empirical study showing that HPs who perceive AI as useful and capable of reducing their workload, are more likely to trust it and use it as a decision-making aid tool in their practice. Conversely, clinicians who believe that AI carries too many risks are less likely to integrate it into their clinical decision-making processes [24]. Building trust is an important element for the successful adoption of AI tools, and this trust may depend largely on whether these tools meet HPs' expectations (i.e., enhance clinical practice, decrease clinicians' workload). If these tools are seen as reliable aids that support rather than hinder their daily tasks, HPs will be more inclined to trust and integrate them into their practice [25].

AI systems' validity can also affect clinicians' trust, and consequently, AI adoption - an important challenge frequently discussed in the literature [26, 27]. AI validity ensures that these systems meet their functional purpose by abiding to necessary standards for safe and just operation, confirming that they undergo appropriate assessments before being put into use [26]. Validating the inputs and outputs of AI systems is crucial to prevent errors caused by data bias, missing information, or poor data quality [28]. Such issues can undermine data validity, potentially posing risks to patients by leading to misdiagnoses or inappropriate treatments. Therefore, continuous oversight of these systems is essential throughout their entire lifecycle to ensure their responsible and ethical implementation.

Lambert et al.'s paper highlights how safety considerations can also strongly influence AI acceptability. Diverse factors can compromise patient's safety in the context of AI, including erroneous outcomes in complex models, biased data, poor data quality, and 'alert sensitivity'- which refers to the risk of clinicians becoming desensitized to frequent AI alerts, especially when many alerts are low-risk, which could lead to wrongfully dismissing high-risk alerts, causing harm to the patient [29]. Excessive reliance on AI in healthcare can jeopardize patient safety by diminishing the vigilance of HPs, who may overly trust AI recommendations at the expense of their clinical judgment. This overreliance increases the likelihood of errors going unnoticed, as practitioners might overlook critical patient-specific nuances that the AI system fails to recognize. Consequently, misinterpretations or delayed responses to inaccuracies can lead to inappropriate diagnoses or treatments, ultimately compromising patient care [30].

Other factors that might influence the pace of AI adoption in clinical settings include clinicians' access to proper training, their age, their willingness to adopt new approaches, the availability and approval of tools by health authorities, and time constraints [31]. Variations in AI adoption among Canadian provinces can be attributed to varying levels of funding allocated for AI implementation. The Canadian government directs a significant portion of its AI implementation funds to British Columbia, Ontario and Quebec [32]. Additionally, healthcare falls under provincial jurisdiction, which influences how each province manages AI implementation in healthcare [33]. This situation can shape clinicians' perceptions of AI's progress and its impact on their practice. Disparities in AI clinical integration may also arise, potentially leading to inequalities in access to advanced technologies and services, disadvantaging stakeholders in regions that are slower to adopt AI in their health systems.

Addressing the gap in AI black box considerations

The black-box issue refers to the challenge of interpreting outcomes generated by complex AI models [16], which can reduce transparency and generate uncertainty for users. While some AI models, such as decision trees, and Bayesian classifiers, employ more transparent algorithms that allow clear explanations of their outcomes, others- particularly deep learning (DL) algorithms- make it harder to understand how certain conclusions are reached [34, 35]. To address this concern, some researchers have proposed post-hoc explanations, using techniques like Local Interpretable Model-agnostic Explanations (LIME) to decipher these complex mechanisms and make AI models more explainable and transparent [36, 37]. Although the black-box issue has been extensively discussed in the literature, particularly as AI systems grow more complex, surprisingly, participants in our study did not express the same level of concern [38–40]. In healthcare, many treatments have historically been implemented despite uncertainty about their mechanisms (i.e., lithium), underscoring the longstanding dilemma of uncertainty and inexplicability in medicine [41–43]. This historical context, coupled with current medical practices -where clinicians are often faced with interpreting findings of uncertain or unknown significance-may account for why HPs in our study did not place as much emphasis on the black-box issue, in contrast to its prominence in the literature. The empirical results from our study revealing a gap between the theoretical discourse surrounding the black-box issue and the weight it is given in practice within the healthcare context, underscores the importance of empirical bioethics in highlighting the differences between theoretical debates and practical realities, to better align ethical frameworks with the actual experiences and priorities of HPs [44].

Consent practices for AI use in healthcare

Participants also expressed concerns about the type of information that should be disclosed to patients regarding the use of AI in their healthcare and how informed consent should be obtained. Some believed that a simple mention of AI use would suffice, while others considered it unnecessary, comparing AI to other medical tools that are not always explicitly disclosed to patients. While there was noclear consensus about AI use disclosure and the type of information that should be provided, the literature suggests that patients need to be informed about data-related aspects of AI, including potential risks such as data misuse in the event of a breach, the impact of data bias on results, and erroneous conclusions from causal outputs [45, 46].

AI training requires large datasets, increasing the demand for patient data, which raises questions about how to effectively obtain consent for data collection. Among the options discussed in the literature is the blanket consent model, where patients are asked to give their consent once for the ongoing use of their data [47]. This approach is less burdensome for both patients and those responsible for obtaining consent, but may raise concerns about data privacy and patient autonomy. Alternatively, seeking consent at each stage of data use could provide stronger safeguards for privacy and autonomy, though it may require more resources [47]. As datasets grow and AI systems become more integrated into healthcare, developing effective strategies for informed consent and patient safety is crucial. HPs stressed the importance of using data in ways that benefit individuals while respecting their consent, aligning with principles of data ownership. Although participants strongly believed that patients should own their medical data, the legal determination of data ownership remains ambiguous, as multiple stakeholders including patients, healthcare professionals, and health institutions- share rights and responsibilities regarding health data [48].

Contributions and Limitations

Numerous studies have examined the challenges associated with the implementation of AI in healthcare by gathering perspectives from a range of stakeholders, including HPs in specific institutions (i.e., hospitals), AI researchers and healthcare information technology experts [18, 49–51]. Our study focuses on the views of Canadian HPs from various medical fields (i.e., family physicians, a neurologist, nurses), provinces, and medical institutions to offer a comprehensive understanding of their insights and concerns related to AI implementation in healthcare.

Future research could expand the sample size by including more HPs from additional provinces, territories and international contexts, as the majority of our participants were from Ontario and Quebec. Further, examining the relationship between HPs' perceptions of AI and factors such as age, years of medical experience, and medical specialty would be valuable for understanding whether these characteristics influence attitudes or behaviors toward AI.

A family physician, for instance, would rely on different AI medical tools in their practice compared to a radiologist or a surgeon. For example, a family physician might primarily use an AI application to transcribe patient conversations or assist in diagnosis [52]. In contrast, a radiologist might use AI for imaging analysis, while a surgeon might use it for critical surgical assistance [53, 54]. Therefore, when discussing AI in healthcare, it is crucial to consider its diverse applications tailored to specific medical contexts, highlighting the need for adaptable guidelines for each use case. Further research could focus on targeted applications, identifying challenges unique to each specific context.

Conclusion

AI is playing an increasing pivotal role in our society, withCanada emerging as a global leader in its implementation across various sectors. In healthcare, while AI promises significant improvements in patient care, our findings reveal that its potential is accompanied by notable concerns. HPs, as key stakeholders in this field, have diverse perspectives on AI. Our interviews reveal varying levels of optimism,

with some professionals enthusiastically embracing AI for its potential to reduce administrative burdens, thereby enhancing their working conditions and improving patient care by reducing waiting times and providing more accurate diagnostics. However, others are more cautious, expressing reservations about whether these technologies have demonstrated enough effectiveness and benefits to outweigh the potential risks. These differing views are influenced by factors such as HPs' knowledge and experience with AI play. Nevertheless, all participants expressed concerns about potential ethical implications, including the accessibility and acceptability of AI tools, the transparency and explainability of complex models, and the secure use of data. Ensuring that HPs fully comprehend and effectively integrate AI into their practices is crucial, as their decision to adopt or reject these technologies can significantly impact patient care. Insufficient knowledge or misuse of AI medical devices may lead to errors or disparities in healthcare delivery, while failing to use AI when it could enhance accuracy can negatively affect patient outcomes. This highlights the need to provide HPs with tailored AI education and clear guidelines, helping to boost their confidence and competence in using AI, ultimately ensuring better patient care.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

SBG and HH designed the study. SBG led the data collection process, with contribution from HH. SBG and HH analyzed the data. SBG drafted the article with input from HH. Both authors critically reviewed the article and approved the final version for publication.

Funding

This study was funded through the Research institutional funds of the Université du Québec at Rimouski.

Ethics Approval and Consent to Participate

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Approval for this study was obtained from the research ethics committee at the Committee of the University of Quebec at Rimouski (#CER-121-976) Quebec, Canada in May 2022. Written informed consent was obtained from all study participants prior to data collection.

References

1. Zhang C, Lu Y (2021) Study on artificial intelligence: The state of the art and future prospects. J Ind Inf Integr 23: 100224.
2. Mondal B (2020) Artificial intelligence: state of the art, Recent Trends Adv Artif Intell Internet Things, p389–425.
3. Loureiro SMC, Guerreiro J, Tussyadiah I (2021) Artificial intelligence in business: State of the art and future research agenda. J Bus Res 129: 911-926.
4. Eli-Chukwu NC (2019) Applications of artificial intelligence in agriculture: A review. Eng Technol Appl Sci Res 9: 4377-4383.
5. Duggal P (2023) Artificial Intelligence – Revolutionizing the Healthcare Industry. Canadian Association of Professionals in Regulatory Affairs,
6. Al-Antari MA (2023) Artificial Intelligence for Medical Diagnostics-Existing and Future AI Technology! Diagnostics (Basel) 13: 688.
7. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. Metabolism 69: S36–S40.
8. Becker A (2019) Artificial intelligence in medicine: What is it doing for us today? Health POLICY Technol 8: 198–205.
9. Kaul V, Enslin S, Gross SA (2020) History of artificial intelligence in medicine. Gastrointest Endosc 92: 807–812.
10. Mukherjee D, Rahemtulla S, Hakim Z (2024) Six ways to make more of AI in Canadian health care. EY,
11. Briganti G, Le Moine O (2020) Artificial intelligence in medicine: today and tomorrow. Front Med 7: 509744.
12. Attard-Frost B, Brandusescu A, Lyons K (2024) The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. Gov Inf Q 41: 101929.
13. Shahed Al-Haque, Marie-Renée B-Lajoie, Erez Eizenman, et al. (2024) The potential benefits of AI for healthcare in Canada. McKinsey & Company,
14. Yu A (2024) These companies are leveraging AI to improve Canada's stressed health care system. The Globe and Mail.
15. Curzon J, Kosa TA, Akalu R, El-khatib (2021) Privacy and artificial intelligence. IEEE Trans Artif Intell 2: 96–108.
16. Von Eschenbach WJ (2021) Transparency and the black box problem: Why we do not trust AI. Philos Technol 34: 1607–1622.
17. Roselli D, Matthews J, Talagala N (2019) Managing bias in AI. In Companion proceedings of the 2019 world wide web conference (pp. 539-544).
18. Catalina QM, Fuster-Casanovas A, Vidal-Alaball J, Escalé-Besa A, Marin-Gomez FX, et al. (2023) Knowledge and perception of primary care healthcare professionals on the use of artificial intelligence as a healthcare tool. Digit Health 9: 20552076231180511.
19. Vaismoradi M, Turunen H, Bondas T (2013) Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 15: 398-405.
20. Bradshaw C, Atkinson S, Doody O (2017) Employing a qualitative description approach in health care research. Glob Qual Nurs Res 4: 2333393617742282.
21. Adeoye-Olatunde OA, Olenik NL (2021) Research and scholarly methods: Semi-structured interviews. J Am Coll Clin Pharm 4: 1358–1367.
22. Campbell KA, Orr E, Durepos P, Nguyen L, Li L, et al. (2021) Reflexive thematic analysis for applied qualitative health research. Qual Rep 26: 2011–2028.
23. Dhakal K (2022) NVivo. J Med Libr Assoc JMLA 110: 270.
24. Shamszare H, Choudhury A (2023) Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. Healthcare (Basel)11: 2308.
25. Hsieh PJ (2023) Determinants of physicians' intention to use AI-assisted diagnosis: An integrated readiness perspective. Comput Hum Behav 147: 107868.
26. Myllyaho L, Raatikainen M, Männistö T, Mikkonen T, Nurminen JK (2021) Systematic literature review of validation methods for AI systems. J Syst Softw 181: 111050.
27. Park SH, Choi J, Byeon JS (2021) Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence. Korean J Radiol 22: 442.
28. Evans H, Snead D (2024) Why do errors arise in artificial intelligence diagnostic tools in histopathology and how can we minimize them? Histopathology 84: 279–287.
29. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, et al. (2023) An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. Npj Digit Med 6: 111.
30. Grote T, Berens P (2022) How competitors become collaborators-Bridging the gap (s) between machine learning algorithms and clinicians. Bioethics 36: 134–142.
31. de Grood C, Raissi A, Kwon Y, Santana MJ (2016) Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 335-344.
32. Brandusescu A (2021) Artificial intelligence policy and funding in Canada: Public investments, private interests. Priv Interes.

33. Allin S (2008) Does equity in healthcare use vary across Canadian provinces? Healthc Policy 3: 83.

34. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H (2019) Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 9: e1312.

35. Rai A (2020) Explainable AI: From black box to glass box. J Acad Mark Sci 48: 137–141.

36. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, et al. (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58: 82-115.

37. Moradi M, Samwald M (2021) Post-hoc explanation of black-box classifiers using confident itemsets. Expert Syst Appl 165: 113941.

38. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). Ieee Access 2018; 6: 52138–52160.

39. Zednik C. Solving the black box problem: A normative framework for explainable artificial intelligence. Philos Technol 2021; 34: 265–288.

40. Wadden JJ (2022) Defining the undefinable: the black box problem in healthcare artificial intelligence. J Med Ethics 48: 764–768.

41. Harish V, Morgado F, Stern AD, Das S (2021) Artificial intelligence and clinical decision making: the new nature of medical uncertainty. Acad Med 96: 31–36.

42. London AJ (2019) Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 49: 15–21.

43. Durán JM, Jongsma KR (2021) Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics 47: 329–335.

44. Solomon MZ (2005) Realizing bioethics' goals in practice: ten ways "is" can help "ought". Hastings Cent Rep 35: 40–47.

45. Kiener M (221) Artificial intelligence in medicine and the disclosure of risks. AI Soc 36: 705–713.

46. Khanna S, Srivastava S (2020) Patient-centric ethical frameworks for privacy, transparency, and bias awareness in deep learning-based medical systems. Appl Res Artif Intell Cloud Comput 3: 16–35.

47. McKeown A, Mourby M, Harrison P, Walker S, Sheehan M, et al. (2021) Ethical issues in consent for the reuse of data in health data platforms. Sci Eng Ethics 27: 1-21.

48. Balthazar P, Harri P, Prater A, Safdar NM (2018) Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. J Am Coll Radiol 15: 580-586.

49. Alami H, Lehoux P, Papoutsi C, Shaw SE, Fleet R, et al. (2024) Understanding the integration of artificial intelligence in healthcare organisations and systems through the NASSS framework: a qualitative study in a leading Canadian academic centre. BMC Health Serv Res 24: 701.

50. Shinners L, Aggar C, Grace S, Smith S (2020) Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: An integrative review. Health Informatics J 26: 1225-1236.

51. Hummelsberger P, Koch TK, Rauh S, Dorn J, Lermer E,et al. (2023) Insights on the current state and future outlook of artificial intelligence in healthcare from expert interviews. JMIR AI 2: e47353.

52. Allen MR, Webb S, Mandvi A, Frieden M, Tai-Seale M, et al. (2024) Navigating the doctor-patient-AI relationship-a mixed-methods study of physician attitudes toward artificial intelligence in primary care. BMC Prim Care 25: 42.

53. Sorantin E, Grasser MG, Hemmelmayr A, Tschauner S, Hrzic F, et al. (2021) The augmented radiologist: artificial intelligence in the practice of radiology. Pediatr Radiol 52: 2074-2086.

54. Wall J, Krummel T (2020) The digital surgeon: How big data, automation, and artificial intelligence will change surgical practice. J Pediatr Surg 55: 47–50.