

Preparing for the future: How organizations can prepare boards, leaders, and risk managers for artificial intelligence

Arun Dixit, MEng, PEng, CHE, LSSBB¹[®]; Jennifer Quaglietta, PEng, MBA, CHE, PMP, LSSGB¹; and Catherine Gaulton, JD, LLM, RN (Retired), CRM, ICD(D)¹



Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/08404704211037995 journals.sagepub.com/home/hmf

Abstract

Artificial Intelligence (AI) is the notion of machines mimicking complex cognitive functions usually associated with humans, such as reasoning, predicting, planning, and problem-solving. With constantly growing repositories of data, improving algorithmic sophistication and faster computing resources, AI is becoming increasingly integrated into everyday use. In healthcare, AI represents an opportunity to increase safety, improve quality, and reduce the burden on increasingly overstretched systems. As applications expand, the need for responsible oversight and governance becomes even more important. Artificial intelligence in the delivery of healthcare carries new opportunities and challenges, including the need for greater transparency, the impact AI tools may have on a larger number of patients and families, and potential biases that may be introduced by the way an AI platform was developed and built. This study provides practical guidance in the development and implementation of AI applications in healthcare, with a focus on risk identification, management, and mitigation.

Introduction

Artificial Intelligence (AI) is the notion of machines mimicking complex cognitive functions such as reasoning, predicting, planning, and problem-solving. With growing data repositories, improving algorithmic sophistication, and faster computing resources, AI is becoming increasingly integrated into everyday use.

In Canada, healthcare spending as a percentage of Gross Domestic Product (GDP) has been rising for decades,¹ and AI represents a potential opportunity to reduce burdens on overstretched healthcare systems. In recent years, progress has advanced to where AI systems can exceed human performance in certain tasks.² In Canada, AI applications are being developed for clinical and non-clinical settings, with examples including AI-enhanced peer review of radiological images³ and tools for improving supply chain management efficiency.⁴ While AI may hold great promise, the actual number of sustained implementations in healthcare settings presently remains limited.⁵ As applications of AI continue to develop, the need for careful risk management becomes increasingly important.

This study proposes a risk management framework intended for use by boards, senior leaders, and risk managers in Canadian healthcare organizations when adopting and implementing AI. *Risk Management in AI* provides an overview of major risk types that can occur with AI in healthcare settings. *Risk Management Guiding Principles* presents risk management guiding principles, and *Applied Framework* presents a framework for the management of risks. The study is intended to support healthcare organizations in establishing their own processes for initiating, planning, prioritizing, overseeing, and governing AI-based projects.

Risk management in AI

The introduction of new technologies, particularly in healthcare settings, requires careful planning and risk management. To support the development of the proposed framework, a keyword search for articles providing guidance and risk management advice for healthcare administrators was conducted on prominent Canadian healthcare journals. Canadian healthcare journals were included due to the unique nature of the Canadian healthcare legislative system and accountabilities. A diagram outlining the search criteria is included in Figure 1. These articles were reviewed to understand the current state of risk management guidance for a Canadian healthcare audience.

Ethical risks

Artificial intelligence applications in healthcare must be compliant with evolving regulations and ethical guidelines. As of the time of writing this article, Canada does not have a regulatory framework specifically for AI applications.⁶ In November 2020, the Office of the Privacy Commissioner of Canada presented recommendations to the Personal Information Protection and Electronic Documents Act (PIPEDA), which are intended to help enable the benefits of AI while maintaining the rights of individuals to privacy.⁷ Recently, the Government of Canada has proposed Bill C-11 which would impact the use of data in automated decision-making.⁸ Canadian healthcare

Arun Dixit, Healthcare Insurance Reciprocal of Canada, North York, Ontario, Canada.

E-mail: adixit@hiroc.com

¹ Healthcare Insurance Reciprocal of Canada, Toronto, Ontario, Canada.

Corresponding author:

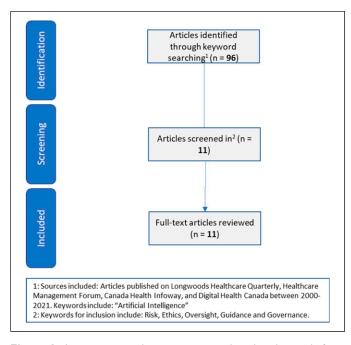


Figure 1. Literature search parameters employed in this study from prominent Canadian journals.

organizations would also be required to be compliant with provincial legislation as appropriate.

Efforts to adopt AI should also address inequities that may create further gaps in the quality of, access to, and delivery of healthcare services. McCradden et al. argue that biases in health data may represent a significant threat to the ethical adoption of AI.⁹ Therefore, a critical step is ensuring a meaningful problem of interest is being solved using appropriate and representative data. Furthermore, while AI systems can perform certain tasks faster and more accurately than humans, health leaders must carefully consider and define the boundaries that would be placed on these systems to protect against the introduction of biases.

Suresh and Guttag proposed a framework for the distinct sources of bias which can arise from the use of AI systems, including¹⁰:

- Historical bias: When the data elements available have evolved over time and this distinction is not made clear during system development.
- 2. **Representation bias:** When datasets do not appropriately or comprehensively reflect the needs and interests of all subgroups of a population.
- 3. **Measurement bias:** When data collection processes have been used inconsistently or inaccurately.
- Aggregation bias: When unique subgroup characteristics are lost, such as identifiers that may be useful for exploring geographical variation in service delivery.
- Evaluation bias: When data used to build a system is semantically different from what would be seen postdeployment.
- 6. **Deployment bias:** When a system is used for performing tasks beyond its original scope of use.

Governance risks

Wiens et al. argue that comprehensive governance and stakeholder engagement is essential to the success of any AI initiative.¹¹ In AI initiatives, stakeholders may include internal and external parties from varied backgrounds and roles. Moreover, stakeholders in healthcare-based AI projects should include representatives of patients and families. Stakeholder groups in AI initiatives include¹¹:

- Knowledge experts: Including clinical experts, ethicists, researchers, information and technology experts, and change management and implementation professionals.
- 2. **Decision-makers:** Including leaders at the local, regional and system level, policy-makers, administrators, and boards.
- End-users: Including patients and families, clinicians, support staff, and other stakeholders impacted by changes.

Successful implementation of AI systems requires collaboration between all three stakeholder groups.

Governance risk can be demonstrated by considering the scenario where an organization is developing an AI system to support clinical decision-making. The development of AI systems generally benefits from larger amounts of data; however, this organization only has a small number of patient records for the problem of interest. The organization could consider using only its internal data, entering into data sharing agreements with other organizations, or creating synthetic data. Synthetic data are created via algorithms rather than being created by actual processes or individuals and may help improve performance of certain AI systems.¹² Moreover, the organization is interested in partnering with AI knowledge experts for the development and deployment of the system. In pursuing this AI application, ownership of each aspect of the solution, including where decision-making, accountability, and liability will reside, must be carefully defined and agreed to.

Performance risks

Effectively building an AI system requires the clear definition of a solution objective. Thomas and Uminsky highlight the importance of choosing appropriate performance metrics and the business risks that may arise from incorrect choices.¹³ Consider an AI application built to interpret radiological images. If, hypothetically, abnormal findings are only present in 1% of the images, a system that always declared no abnormal findings would be wrong 1% of the time and would yield a misleading accuracy of 99%. Moreover, despite significant advances in recent years, no AI model presently produces perfect results.¹⁴ Artificial intelligence systems may produce false positive predictions and false negative predictions, and health leaders must carefully consider these impacts what performance thresholds are tolerable. Forcier, Khoury, and Vezina present three scenarios where organizations could be liable stemming from the performance of an AI system. The first is when damages are claimed by a patient against the company that created the AI system, the second is when damages are claimed by a patient against a hospital or healthcare organization, and the third is when damages are claimed by the physician or healthcare organization against the company that created the system. Moving forward, organizations will need to be aware of evolving Canadian case and civil law and their potential impacts to liability in the use of AI systems in all scenarios above.¹⁵

Recently, deep learning has emerged as a technique capable of processing large volumes of data to generate predictions. However, these systems may contain millions of parameters that govern how they function, and it can be difficult to trace how a particular decision was made. Interpretability of AI systems and the ability to clearly explain and trace its decisions is critical for enabling their adoption, as well as for their continuous improvement both pre- and post-deployment.¹⁶

Implementation risks

Since the 1960s, thousands of AI models have been developed; however, very few have been implemented in practice. Utsun argues that this may be due to **disuse** and **misuse** of AI.¹⁷

- Disuse: When clinicians or end-users do not trust how a model was built or may not have been involved in the development process.
- Misuse: When a system was not built using appropriate data, it has inaccuracies, or it was created with unintentional biases.

Interactions between humans and AI systems must also be carefully considered. The intent may be to create more time between clinicians and patients, but AI systems could lead to unintended consequences such as increased time spent between humans and computers. Greater involvement of end-users and patients and families is a key enabler of the viability of AI systems.

In 2021, the European Commission published the *Proposal* for a Regulation on a European Approach for AI.¹⁸ This document outlines three categories of AI systems based on their overall risk profile. These categories include

- Unacceptable risk systems: Applications that pose a clear threat to an individual's security or fundamental rights, including systems that can cause physical or psychological harm.
- High-risk systems: Applications that may impact critical processes or functions, including critical infrastructure or other essential services.
- 3. Low-risk systems: Applications that do not pose a material threat to health or safety, such as spam filters or chatbots.

Organizations adopting AI would need to consider which category a proposed application belongs to and consult appropriate with subject matter experts and stakeholders to determine satisfactory performance thresholds or boundary conditions in which the system would operate.

Security risks

Artificial intelligence has the capacity to impact many patients and may have been built using numerous data sources. Therefore, careful consideration must be given to protecting these systems against vulnerabilities, cyberattacks, and unauthorized access while maintaining the integrity and confidentiality of personal health information.

Artificial intelligence systems are highly dependent on the availability of high-quality, reliable datasets. Even slight perturbations to datasets provided to AI systems can significantly alter their predictions or recommendations. In a controlled environment, Jiawei et al. demonstrated that modifying a single pixel on images ingested into to an image recognition system greatly altered what it believed it was seeing.¹⁹ This example demonstrates the critical need for strong cybersecurity strategies to protect from external threats and breaches.

While AI has the potential to produce innovative improvements, a careful and deliberate assessment of security risks must be taken prior to the start of each new project and updated throughout the lifecycle of the initiative. Organizations can seek information and guidance cyber security measures from several organizations including the Canadian Centre for Cyber Security, which has issued over 2,000 resources since 1998.²⁰

Risk management guiding principles

The principles presented here are intended to support health leaders in risk management from concept development to implementation and monitoring of AI systems. The research team developed this list by using the Osborn method to identify as many recommended actions as possible. This method involves identifying as many possible answers to a question of interest. The list is then validated against any known standards or guidance documents and refined to remove any redundant items. The team consulted existing risk management guidance documents for validating the list of guiding principles proposed in this study.^{18,21-23}

Clearly define the value proposition of AI systems

- 1. Consult widely with stakeholders, including clinicians, patients, and families to develop meaningful questions to be answered.
- Consider applications aligning with one or more dimensions of quality, which include accessibility, appropriateness, effectiveness, efficiency, equity, integration, patient-centredness, a population healthfocus, and safety.²⁴
- 3. Define problems based on areas of need and then identify data requirements, as opposed to selecting problems based on available data.

Guiding Principle	Key Questions: Boards of Directors	Key Questions: Health Leaders	Key Questions: Risk Managers
Ethical risks: Clearly define the value proposition of AI systems	 What services, care pathways, or client populations have the highest burden and need? Can Al help create fair, equitable access? Could an Al system introduce new biases or inequities in the healthcare system? What boundaries should be placed on Al systems? What types of decisions should they be allowed to make? 	 How can AI help address potential gaps at our organization? Have these applications been proven or is it a novel application? What boundaries should be placed on AI systems? What decisions should they be allowed to make? Does the application of AI lead to improvements that are only possible at infeasible costs? 	 Would an AI solution that makes errors be tolerable to deploy in practice? What if an AI system makes errors, but does so less frequently than humans? Does our organization have appropriate datasets with reliable means of capturing inequities and diversity? Has the proposed AI solution and objective undergone an independent ethics review?
Governance risks: Establish comprehensive governance and oversight	 Who in the organization has ultimate decision-making authority in relation to Al initiatives and what is the framework for such decisions? Are patients and families included? If vendors or other external parties are involved in developing a new solution, who owns the solution? Who is accountable? Any Al solution must be designed carefully around the workflows in which it will be used. Are there unintended consequences of its use? 	and have these processes and required approvals been communicated and enforced in policies?What consideration is given to the future of work, including how AI may impact workflows and functions?	 How are Al-related decisions made and have these processes and required approvals been communicated and enforced in policies? Who is accountable for recommendations made by Al systems in our organization? Are all stakeholders impacted by the outcomes of an Al project included in governance and oversight? Have any stakeholders been missed?
Performance risks: Apply rigorous methods in building AI systems	 No Al model will produce perfect results. What performance threshold is acceptable? How is this determined, and who is accountable? Is a solution developed with in- house resources and expertise, or are external partners involved? How are the external partners chosen? How are their skills and services validated? What support do any external providers offer post-deployment, and what is done to ensure accountability? 	 Is a solution developed with in-house resources and expertise, or are external partners involved? How are they chosen and validated? What support do any external providers offer post-deployment, and what is done to ensure accountability? Al systems often require vast amounts of data to train and build. Are there scenarios where the organization considers partnerships with other institutions to augment datasets? 	 AI systems may produce false positive or false negative predictions. Is one error type worse than others? What error rate is tolerable, and how is this determined? Is the data used to build and train the system representative of what it will see in real life? How does an AI system continue to learn post-deployment? What is the ongoing process to collect for continuous improvement?
Implementation risks: Apply change management tools and processes	 Who is ultimately accountable for recommendations made by an AI system? If an AI system provides a recommendation that conflicts with the advice of a clinician, who makes the final decision? Have patients provided consent for their data to be used for development of an AI system? Have clients provided consent to have their treatment informed or guided by an AI system? 	it provides in a way that a diverse user population can understand?	 How is the AI system making recommendations or acting? What inputs does the system consider, and how are they evaluated? How do users provide feedback if they think the output of an AI model is incorrect? How is this feedback used to update the model? During incident reviews and investigations, are the datasets used to train an AI system also disclosed? If so, how?

Table I Kay avastians to	a a mai da mu u da a ma a a sta bili a bim	lining and		a na sana ant in baalth sana an sainatiana
I able I. Key questions to	consider when establishin	g policies and	procedures for risk m	nanagement in healthcare organizations

(continued)

Guiding Principle	Key Questions: Boards of Directors	Key Questions: Health Leaders	Key Questions: Risk Managers
Security risks: Create strict privacy and security protocols	I. What controls are in place to manage risk of privacy breaches or other cyber security incidents? How are data integrity, privacy, and security maintained?	I. What controls are in place to manage risk of privacy breaches or other cybersecurity incidents? How are data integrity, privacy, and security maintained?	 I. Al models may also have been built using data from many sources. Where are these data stored, and how is access to this maintained? What measures have been put in
	2. What measures have been put in place to prevent hardware and software faults that could result in data being compromised?	2. What measures have been put in place to prevent hardware and software faults that could result in data being compromised?	place to ensure that data are secure?Is there a way to know if data have been corrupted and how?3. Do additional protocols or mitigation
	3. What is the business continuity plan in the event of a service interruption of an AI system?	<i>,</i> ,	e 1 1

Table I. (continued)

Abbreviation: AI, artificial intelligence.

- 4. Obtain feedback from independent stakeholders on the utility of proposed applications.
- 5. Consider frameworks including the Learning Health System for sustainably transforming processes with data and knowledge.²⁵

Establish comprehensive governance and oversight

- 1. Consider creating a standalone AI Steering Committee with oversight on initiation, planning, execution, and monitoring of AI projects.
- 2. Steering Committee membership should include but not be limited to clinicians and other end-users, patient and family advisors, ethicists, risk managers, policy-makers, administrative leaders, and information technology professionals.
- 3. Carefully evaluate user-system interactions to understand where processes and functions may change post-deployment.
- 4. Apply project management tools including governance charts, terms of reference, and accountability agreements for all AI initiatives.
- 5. Conduct assessments to identify potential unintended consequences from the use of AI systems.

Apply rigorous methods in building AI systems

- Establish minimum data quality specifications for AI solutions. These specifications should consider accuracy, completeness, consistency, credibility, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability, and recoverability.²⁶
- 2. Define comprehensive use cases and acceptance plans pre-implementation.
- 3. Pre-deployment, conduct validation trials with clinicians, experts, and other end-user groups.
- 4. Create monitoring plans including outcome, process and balancing metrics to ensure the system is performing as intended post-deployment.

5. Develop strategies for the tracking and analysis of errors, near misses, and overrides post-deployment.

Apply change management tools and processes

- 1. Develop comprehensive communication plans for stakeholder engagement.
- 2. Create a training and communication plan outlining how the system functions, makes decision, and generates recommendations or decisions.
- 3. Develop feedback loops for monitoring performance and usability of any solution.
- 4. Created dedicated functions to act on feedback loops to implement post-implementation improvements.
- 5. Develop escalation plans and continuously update them via post-deployment feedback.

Create strict privacy and security protocols

- 1. Establish data security plans for AI initiatives, which at a minimum include:
 - a. An inventory of data assets.
 - b. Access permission and controls.
 - c. Computing software and hardware controls.
 - d. Protocols for transmitting, storing, and accessing data.
 - e. Data retention and destruction processes.
- 2. Deliver frequent training and communication, focussing on threat identification and response plans.
- 3. Establish data sharing and access agreements which document data access, usage, and ownership policies.
- 4. Define business continuity and disaster recovery plans for the AI system, including end-to-end infrastructure and resiliency controls.
- 5. Conduct regular simulations to ensure the appropriateness of continuity and recover plans.

While AI holds potential to improve healthcare systems, its realization is dependent on the careful application of risk

management principles to ensure sustainable and effective implementations.

Applied framework

Table 1 intends to support leaders in developing appropriate policies and procedures for risk management in their organizations. These questions were identified by the research team using the Osborn method with the goal of creating a concise checklist to support organizational oversight and risk management.

Conclusion

Artificial intelligence is a complex activity that requires careful risk management and oversight, particularly in Canadian healthcare organizations. While AI holds potential to improve many aspects of healthcare service delivery, risks must be carefully mitigated to prevent unintended consequences. A combination of administrative and technological solutions must be employed by healthcare organizations-and even when those steps are employed, organizations must remain vigilant about keeping protective measures current and viable. In this study, we provide a summary of major risks and present a framework to serve as a starting point for risk management in the adoption and implementation of AI in Canadian healthcare organizations. The utility of this work is in supporting boards, senior leaders, and risk managers to develop appropriate internal processes and controls for managing risk in AI initiatives in which they participate.

Authors' note

The authors of this study are employed by the Healthcare Insurance Reciprocal of Canada (HIROC). None of the authors are receiving funding or benefits related to the submission of this study.

ORCID iD

Arun Dixit () https://orcid.org/0000-0001-9562-4593

References

- 1. Canadian Institute for Health Information. *National Health Expenditure Trends, 1975 to 2019.* Ottawa: CIHI; 2019.
- 2. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* 2015;115: 211-252.
- University Health Network. Fostering continuous improvement in diagnostics. 13 November 2019. [Online]. Available at: https:// www.uhn.ca/corporate/News/Pages/Fostering_continuous_ improvement in diagnostics.aspx. Accessed May 10, 2021.
- Benzidia S, Makaoui N, Bentahar O. The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance, *Technol Forecast Soc Change*. 2021;165:120557.
- Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.

- Shah R, McGreevey M, Reynolds M. The future of AI regulation in Canada: what we can learn from the E.U.'s proposed AI framework. Torys LLP; May 25, 2021. [Online]. Available: https://www.torys.com/insights/publications/2021/05/the-futureof-ai-regulation-in-canada. Accessed May 10, 2021.
- Office of the Privacy Commissioner of Canada. A regulatory framework for AI: recommendations for PIPEDA reform. November 2020. [Online]. Available: https://www.priv.gc.ca/en/ about-the-opc/what-we-do/consultations/completed-consultations/ consultation-ai/reg-fw_202011/.
- Minister of Innovation, Science and Industry. Bill C-11, an act to enact the consumer privacy protection act and the personal information and data protection tribunal act and to make consequential and related amendments to other Acts; 2020. https://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c11.html. Accessed May 10, 2021.
- McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health*. 2020;2(5):e221-e223.
- Suresh H, Guttag JV. A framework for understanding unintended consequences of machine learning. 2019. https://arxiv.org/abs/ 1901.10002
- Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* 2019;25(9): 1337-1340.
- Jaipuria N, Zhang X, Bhasin R, et al. Deflating dataset bias using synthetic data augmentation. Paper presented at: Conference on Computer Vision and Pattern Recognition. June 14-19, 2020; Seattle, WA.
- Thomas R, Uminsky D. The problem with metrics is a fundamental problem for AI. arXiv; 2020. https://arxiv.org/abs/2002.08512. Accessed May 10, 2021.
- Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inf Decis Making*. 2020;20(1):310.
- 15. Forcier MB, Khoury L, Vezina N. The use of artificial intelligence in health care: liability issues. *CIRANO Working Papers*; 2020.
- Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? February 2019. [Online]. Available at: https://journalofethics.ama-assn.org/ article/can-ai-help-reduce-disparities-general-medical-andmental-health-care/2019-02. Accessed May 10, 2021.
- Utsun B. Designing for the last mile in machine learning. Toronto, 2020. https://www.cs.umd.edu/event/2020/04/designing-last-milemachine-learning. Accessed May 10, 2021.
- 18. European Commission. Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending centain. Brussels: Union Legistative Acts; 2021.
- Su J, Vargas DV, Sakurai K. One pixel attack for fooling deep neural networks. *IEEE Trans Evol Comput.* 2019;23(5):828-841.
- Canadian Centre for Cyber Security. Information & guidance. April 21, 2021. [Online]. Available: https://cyber.gc.ca/en/ information-guidance.

- Government of Canada. Responsible use of artificial intelligence (AI). May 14, 2021. [Online]. Available: https://www.canada. ca/en/government/system/digital-government/digital-governmentinnovations/responsible-use-ai.html#toc1. Accessed May 10, 2021.
- 22. Government of Ontario. Artificial Intelligence (AI) guidance. Queen's Printer for Ontario. March 30, 2021. [Online]. Available: https://www.ontario.ca/page/artificial-intelligence-ai-guidance. Accessed May 10, 2021.
- World Health Organization. WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use. June 28, 2021. [Online]. Available: https://www.

who.int/news/item/28-06-2021-who-issues-first-global-report-onai-in-health-and-six-guiding-principles-for-its-design-and-use. Accessed May 10, 2021.

- 24. ECFAA. *Excellent Care for All Act.* Ontario: Queen's Printer for Ontario; 2010.
- Institute of Medicine (US). Institute of Medicine (US) Roundtable on Evidence-Based Medicine. Washington: National Academies Press (US); 2007.
- ISO/IEC. Software engineering Software product quality requirements and evaluation (SQuaRE) — Data quality model. ISO/IEC; 2008.