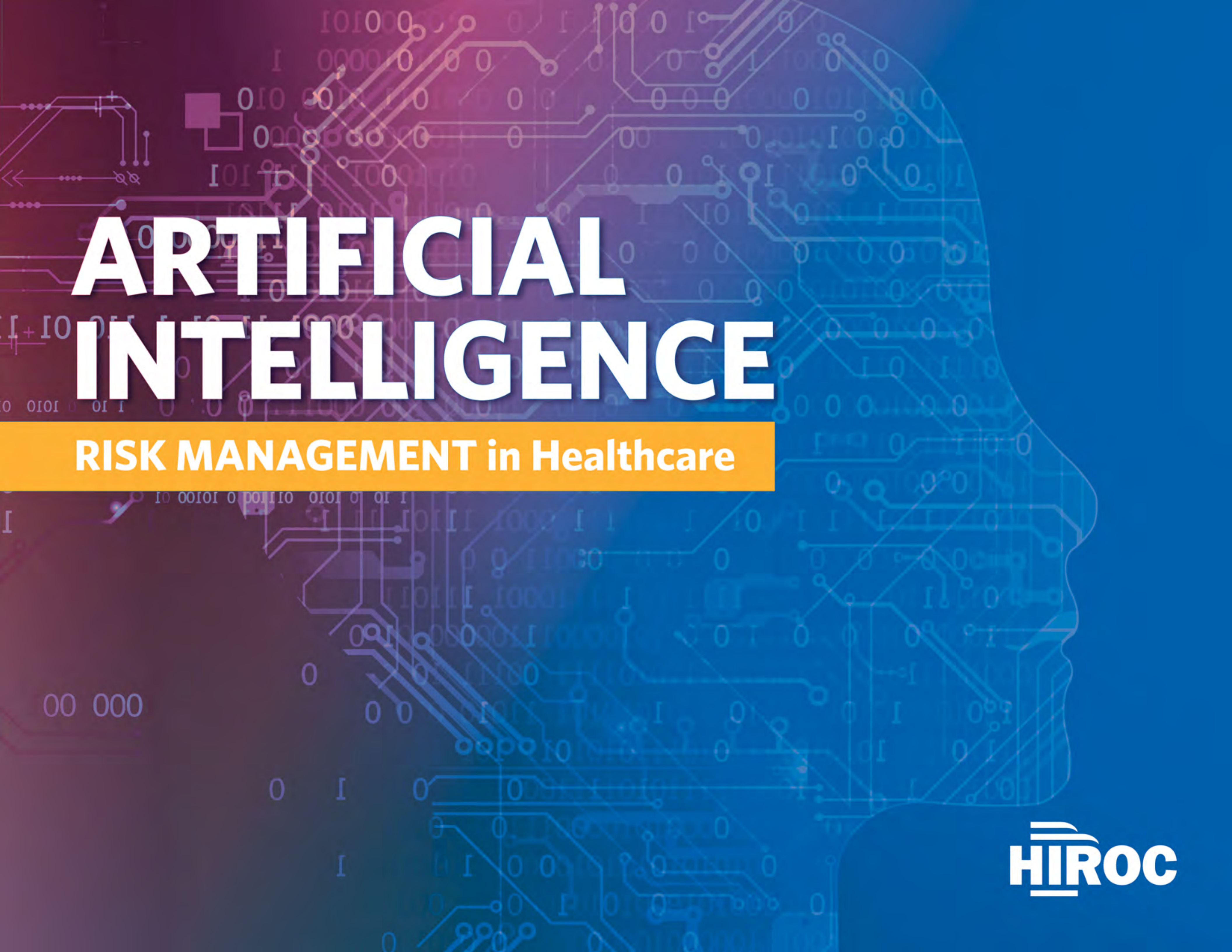


ARTIFICIAL INTELLIGENCE

RISK MANAGEMENT in Healthcare

HIROC

Contents

August 2020

DISCLAIMER: Disclaimer/Terms of Use: This is a resource for quality assurance and risk management purposes and is not intended to provide legal or medical advice. Every effort has been made to ensure that the information is accurate at time of publication.

Healthcare Insurance Reciprocal of Canada | **HIROC.com**

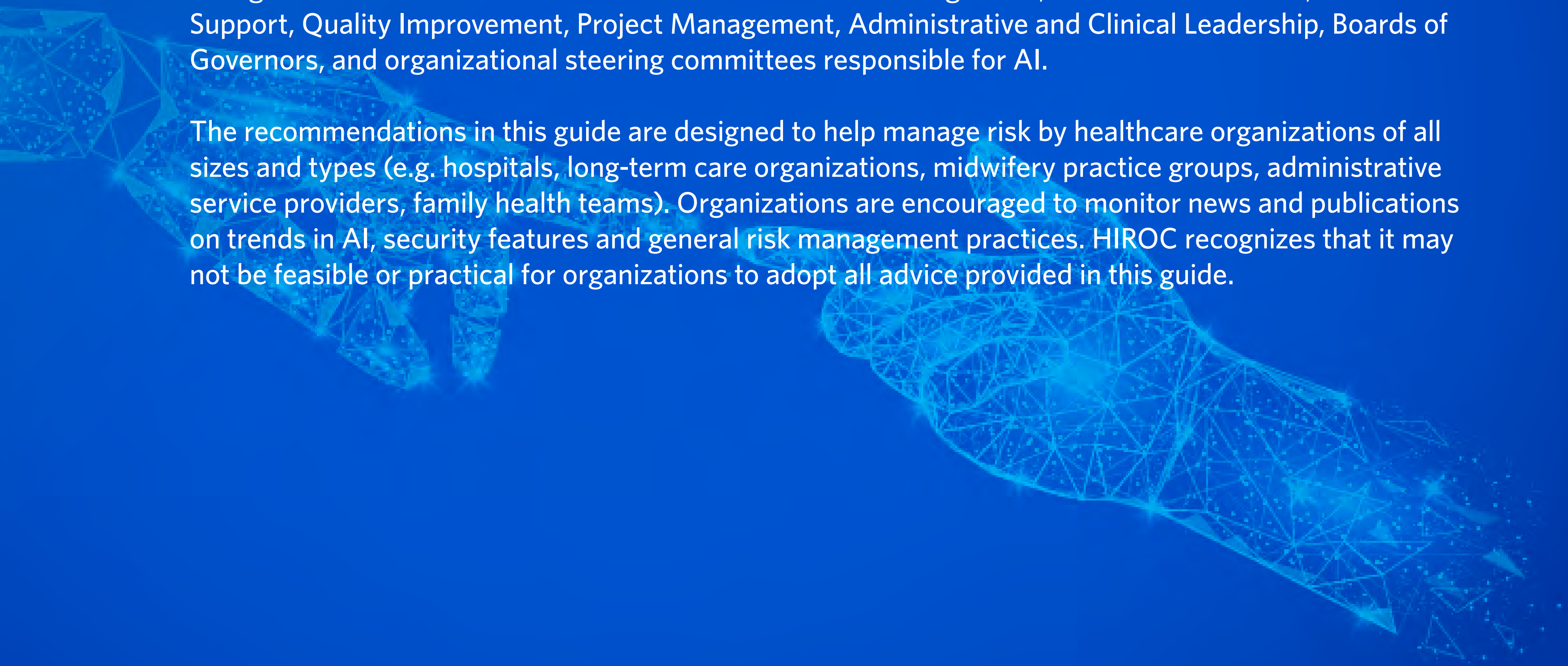
The purpose of this document is to provide practical guidance in the development and implementation of AI applications in healthcare. Our focus is on the identification, management and mitigation of risks. The intent is for this document to support healthcare organizations in establishing their own processes for identifying, initiating, prioritizing, overseeing and governing AI-based projects in which they participate.

Who this guide is intended for

This guide was developed using HIROC's expertise and findings from Canadian and international healthcare-based examples. This guide complements existing resources and frameworks by providing risk management advice for the Canadian healthcare sector.

This guide is intended for individuals in the roles of Risk Management, Information Services, Decision Support, Quality Improvement, Project Management, Administrative and Clinical Leadership, Boards of Governors, and organizational steering committees responsible for AI.

The recommendations in this guide are designed to help manage risk by healthcare organizations of all sizes and types (e.g. hospitals, long-term care organizations, midwifery practice groups, administrative service providers, family health teams). Organizations are encouraged to monitor news and publications on trends in AI, security features and general risk management practices. HIROC recognizes that it may not be feasible or practical for organizations to adopt all advice provided in this guide.



> “AI errors are potentially different for at least two reasons. First, patients and providers may react differently to injuries resulting from software than from human error. Second, if AI systems become widespread, an underlying problem in one AI system might result in injuries to thousands of patients —rather than the limited number of patients injured by any single provider's error."
>
> *(Nicholson-Price, 2019)*

In healthcare, AI offers significant promise to improve complex systems to improve population health, enhance client experience and outcomes, improve staff experience and reduce costs. As applications of AI continue to develop and expand, the need for responsible oversight and governance becomes increasingly important.

Applying AI in the delivery of healthcare introduces new challenges, including the need for transparency in how AI applications are built, the impact AI tools may have on a larger number of clients, and potential biases that may be introduced by the way an AI platform was developed and built.

Healthcare organizations and independent practitioners may face many challenges when adopting or implementing AI-based solutions, including:

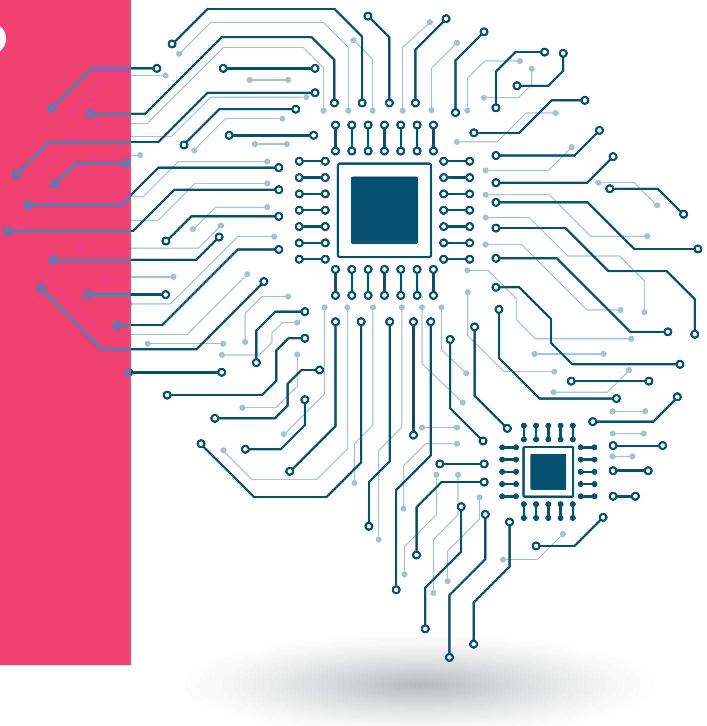
- ✓ Defining appropriate problems to solve using AI
- ✓ Continuously evolving technology
- ✓ A large and diverse range of healthcare users – including patients, families, clients, caregivers, staff, physicians, volunteers and more.

The purpose of this document is to provide practical guidance in the development and implementation of AI applications in healthcare. Our focus is on the identification, management and mitigation of risks. The intent is for this document to support healthcare organizations in establishing their own processes for identifying, initiating, prioritizing, overseeing and governing AI-based projects in which they participate.

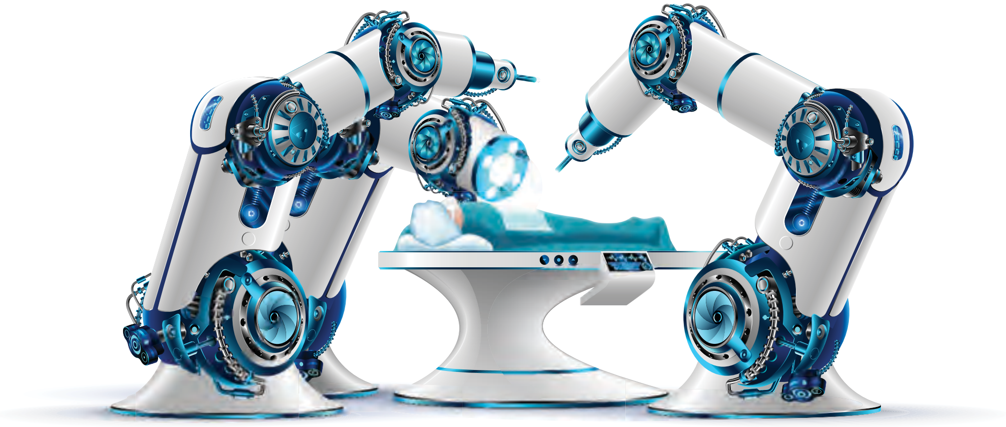
What is Artificial Intelligence?

Artificial Intelligence (AI) is the notion of machines exhibiting and mimicking cognitive functions that are usually associated with humans, such as learning, reasoning, predicting, planning, recognizing, and problem solving. With constantly growing repositories of data, improving sophistication of algorithms and faster computing resources, AI is becoming increasingly integrated into everyday use.

AI methods include a wide variety of tools, such as voice recognition, natural language processing, computer vision, robotics, scheduling, and planning to allow machines to reason, engage and learn with their environment. AI models and applications are frequently built using extremely large datasets, often including millions of data points or more.



Artificial Intelligence in Healthcare



In Canada, healthcare spending as a percentage of Gross Domestic Product (GDP) has been constantly rising for decades (Canadian Institute for Health Information, 2019), and the applications of AI tools present an opportunity to increase safety, improve quality and reduce the burden on increasingly overstretched healthcare systems.

In recent years, progress has advanced rapidly, to the point where AI systems are able to exceed human performance in certain tasks (Russakovsky, Deng, Su, Krause, & Satheesh, 2015). This progress has introduced a shift from research work to demonstrate AI as a proof-of-concept to the implementation of clinically viable applications and their societal impacts.

> **The introduction of new technologies, particularly those implemented in healthcare systems, require careful planning and foresight. Distinct from post-damage and pre-damage control, the precautionary principle is a strategy to cope with uncertainties in the assessment and management of risks. The precautionary principle states that when activities may lead to morally unacceptable harm, actions shall be taken to avoid or diminish that harm."**
>
> *United Nations Educational, Scientific and Cultural Organization, 2005.*

Risks of AI in Healthcare

Several types of risks may arise from the development or implementation of AI systems in an organization. These may include:

Ethical Risks:

Are the right questions being asked, for the right purpose, and do the answers lead to meaningful, value-add improvement?

Case Example:
A research team developed an early warning system to notify clinicians when patients admitted in hospital were at risk of developing sepsis. The model was built using data at two large academic hospitals in the United States and was shown to have good predictive value. During an initial trial, feedback from the clinicians was that the early warning system did not provide them value because the alerts were being triggered when the clinicians already suspected a patient might develop sepsis. Moreover, adding additional non value-added alerts may lead to alarm fatigue by clinicians. (Ginestra, et al., 2019).

Performance Risks:

Are the appropriate AI tools being deployed? How is their success measured and monitored? How and when are corrective actions put in place?

Case Example:
An AI-based decision support tool used by Surgeons to triage patient referrals might be accurate when it is initially deployed, but external factors such as patient populations or clinical guidelines may change over time and the tool may not be 'aware' of these updates. How is the decision support tool updated? How is it continually tested to ensure it is providing recommendations in line with evidence and best-practices? (Beyene, Welemariam, Persson, & Lavesson, 2015)

Governance Risks:

Are all appropriate stakeholders identified and have they been engaged at the right time and in the right way?

Case Example:
In the 1990's a machine learning model was built to predict which patients admitted to hospital were at highest risk of death when diagnosed with pneumonia. The model incorrectly concluded that admitted patients with a prior history of asthma had a lower risk of dying than those without. The model learned this pattern because pneumonia patients with history of asthma were admitted directly to the ICU, and therefore received more aggressive treatment, leading to better outcomes (Caruana, et al., 2015). Without involvement of clinical stakeholders, this error might not have been caught. This example shows how crucial it is to involve the right human experts and established science to prevent unintended consequences and decrease the risk of harm.

Communication Risks:

How does an AI model make decisions and how is this communicated to users with a vast range of expertise and needs?

Case Example:
Deep Learning is an advanced AI technique that is capable of processing and interpreting extremely large volumes of data to generate predictions or automate processes. However, deep learning models may have millions, or even billions of parameters that define how the model functions and makes predictions. With this many parameters, there is a risk of treating an AI system as a "black box" for making predictions. There is a strong need to clearly translate how a deep learning model, or any other AI model processes data to generate recommendations (Denil, Shakibi, Dinh, Ranzato, & de Freitas, 2013).

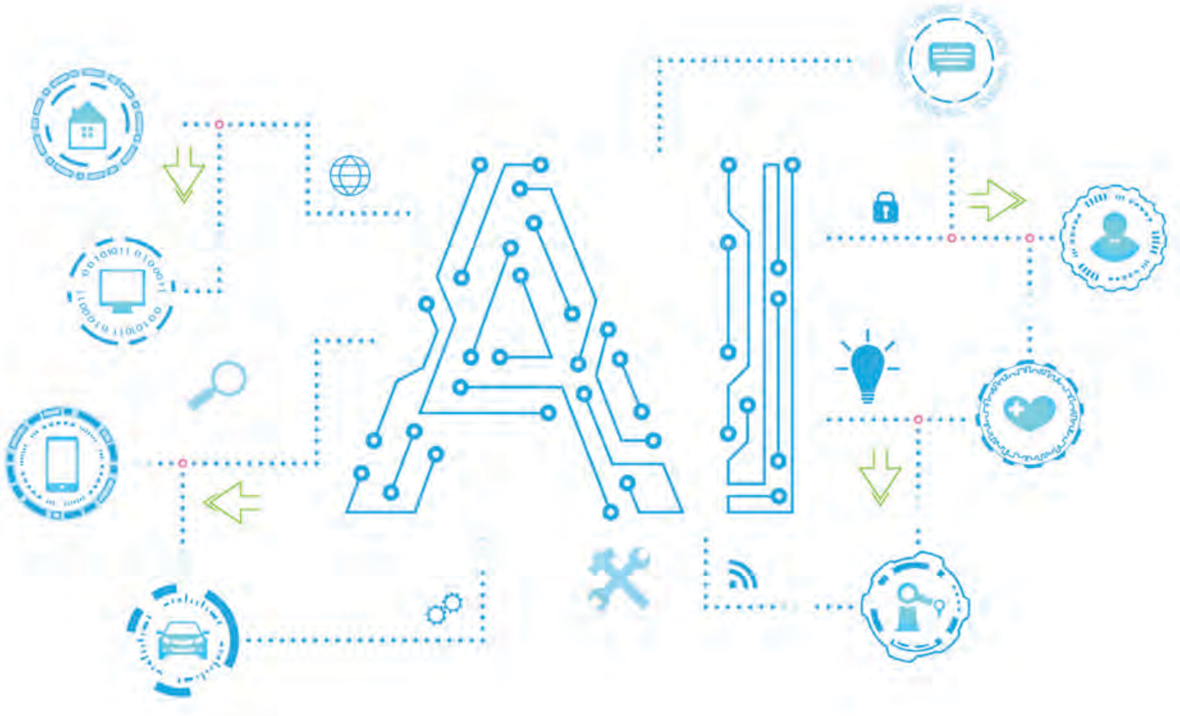
Security Risks:

These include privacy issues, information protection and security risks, data access management and control, other cyber security risks.

Case Example:
Beyond the threat of data breaches and hacks is the threat of compromise to an AI system. Using sophisticated AI techniques, researchers were able to generate fake computed tomography (CT) images that were so convincing that experienced radiologists were not able to tell which images were from actual patients. This case example highlights the need for strong cyber security protocols to be in place to mitigate risks, as a compromised AI system could potentially lead to harm to a large number of patients (Mirsky & Mahler, 2019).

> **Risk-based approaches should be a part of assessing whether an AI or machine learning model should be used for making certain decisions, or for determining what additional controls need to be in place."**
>
> *(Crowe, 2019)*

While AI has the potential to produce many exciting and innovative improvements, a careful and deliberate assessment of the risks must be taken prior to the start of each new project. The following section provides guiding principles to support organizations with identifying appropriate control mechanisms to put in place to mitigate risk.



Guiding Principles

This section presents guiding principles to support organizations with the oversight and governance of AI applications. These guiding principles are not intended to recommend specific areas of focus or clinical applications of AI; rather, they are intended to provide organizations with a framework for developing and managing AI strategies in alignment with their organizational objectives.

These guiding principles are meant to help provide direction and support organizations in conducting risk assessments when developing or deploying AI systems. Prior to any new initiative, at a minimum, the following questions must be considered:

Is AI being used to solve the right problems?

Are the right people involved to solve this problem?

Is the right approach being used to solve this problem?

How does AI work and how do we know it's solving the problem?

How are we protecting privacy and system integrity?

VALUE PROPOSITION

Is AI being used to solve the right problems?

Health inequities refer to inequalities in health that are caused by socially influenceable factors, such as barriers in accessing and receiving healthcare (Public Health Agency of Canada, 2008). Efforts to adopt AI must address inequities that may give rise to further gaps in access and delivery of healthcare services. Applications of AI require consideration for the potential to benefit all users of a healthcare service and not solely for selected populations or subgroups. Furthermore, applications of AI in healthcare should be not only compliant with relevant regulations but must also be ethical. Efforts to implement AI should take deliberate and concerted steps to ensure that biases in AI capabilities are avoided or minimized.

More generally, what boundaries should be placed on AI systems? Would an AI solution that makes errors be tolerable to deploy in practice? What if the AI system still makes errors, but does so less frequently than human operators? Should AI systems be allowed to make life-or-death decisions?

Structured datasets are often used in the development of AI applications because they require less post-hoc pre-processing and standardization. As an example, International Classification of Disease (ICD) codes are included as part of the documentation in a patient's electronic health record. ICD codes are used primarily for billing purposes, and there may be discrepancies between these codes and the patient's complete diagnosis. ICD codes are regularly used in AI and Machine Learning studies, particularly those interested in predicting disease trajectories and outcomes (Wiens, et al., 2019). Moreover, the 9th revision includes over 12,000 diagnostic codes and assigning the correct codes is dependent on the amount and quality of information provided by the physician (O'Malley, et al., 2005). Errors or omissions in the assignment of the ICD codes may lead to data quality issues that may lead to incorrect findings when compared to actual patient data.

AI systems can perform many types of tasks faster and more accurately than humans, especially large, computational tasks. AI systems, however, are less capable of identifying emotional cues. The design of AI systems should therefore build upon the complementary strengths of humans and machines. As an example, if an AI algorithm generates many alerts that are false positives or false negatives, it may lead to disuse or mistrust of the system by users over time.

In addition, biases could be introduced if an AI algorithm is not trained on data that is appropriate or relevant to what the algorithm will see once deployed. A recent study (Suresh & Guttag, 2019) identified several types of biases that could influence the validity of AI models, including:

- **HISTORICAL BIAS:** When a model is trained on data that is not fully representative of a problem or task. As an example, this may occur when data on an individual's income, social status, or education is missing from an analysis of health inequalities.
- **REPRESENTATION BIAS:** This occurs when datasets do not include all relevant and appropriate subgroups of a population.
- **MEASUREMENT BIAS:** This bias occurs if data is not available to the appropriate level of precision. For example, age may be a relevant data element, but might only be available by binned age ranges.
- **AGGREGATION BIAS:** This occurs if distinct populations are combined and data granularity is lost. This may occur, for example, if all client records from a province were combined and regional identifiers that may be useful for exploring geographical variation in healthcare service delivery were lost.
- **EVALUATION BIAS:** This type of bias may arise if the model used to build an AI model is different from what it sees in deployment. As an example, if a model for use in an Emergency Department was built using data from April to September, it may not succeed in applications during winter surge periods.
- **DEPLOYMENT BIAS:** This can occur if a model is used for purposes beyond what it was originally built for. For example, an application for determining abnormal findings in MRI images of brains would not be appropriate for use on images of lungs.

KEY QUESTIONS TO CONSIDER:

1. What are the populations with the highest burden and need?
2. Is there a gap for this population that is amenable for improvement? How precisely can AI help close this gap?
3. Is there a willing group of healthcare providers to test and implement AI to address these gaps?
4. Do available datasets include reliable means of capturing inequities and diversity? How accurate is the data being used to build an AI model?
5. Does the application of AI lead to improvements that are only possible at infeasible costs?
6. Could an AI system introduce new biases or inequities in the healthcare system?
7. Would the development of an AI model lead to further inequities or access barriers?
8. How will the AI system be used in practice? What interactions do human users have with this system?
9. What decisions or recommendations is the AI system allowed to make? Are there boundaries placed on its decision-making capabilities?
10. Who in our organization has ultimate decision-making authority in relation to AI initiatives and what is the framework for such decisions?
11. Has the proposed AI solution and objective undergone an independent ethics review?

GOVERNANCE:

Are the right people involved to solve this problem?

Projects of any type are rarely successful without considerable stakeholder engagement. Projects may include internal and external stakeholders from diverse backgrounds and roles. Moreover, stakeholders in healthcare-based AI projects should include representatives of clients and families. In the context of AI applications in healthcare, three stakeholder categories are equally critical for projects to be successful (Wiens, et al., 2019):

✓ KNOWLEDGE EXPERTS

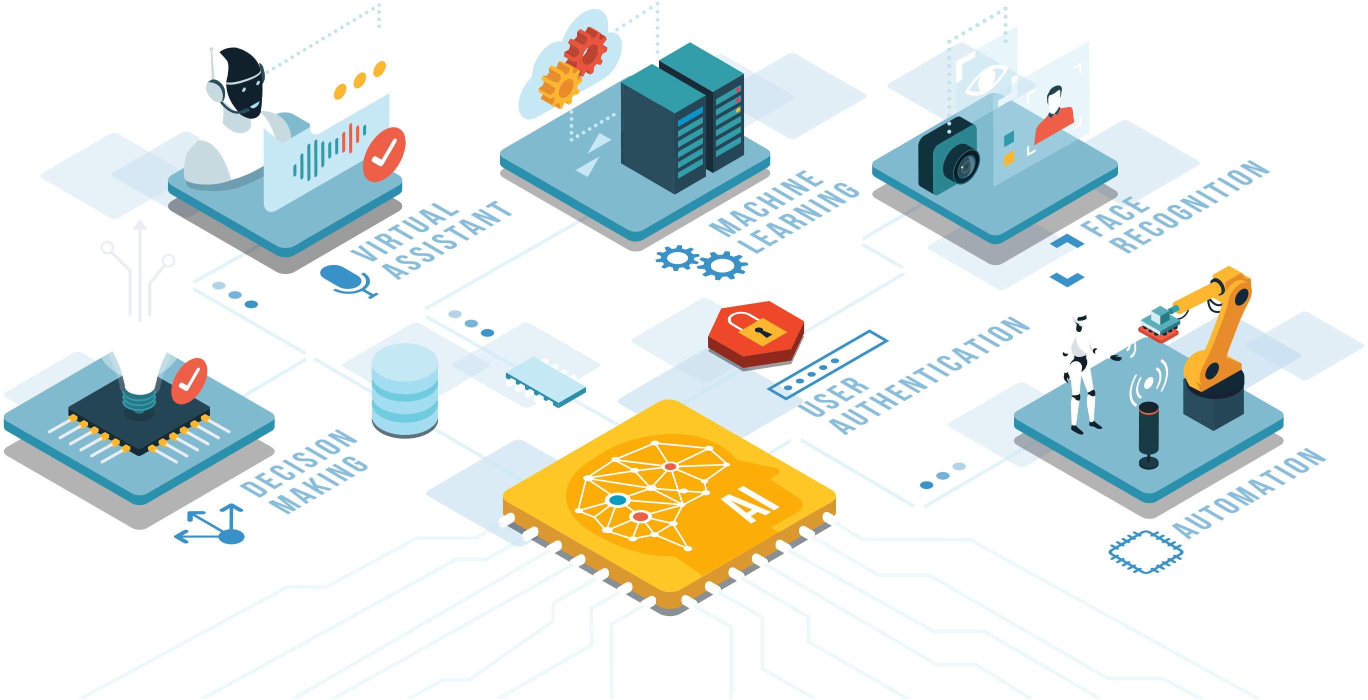
These include clinical experts, ethicists, machine learning and AI researchers, health information and technology experts and experts in change management and implementation.

✓ DECISION-MAKERS

These include hospital administrators, institutional leadership and regulatory agencies.

✓ USERS

These include clients and families, clinicians, support staff and other stakeholders impacted by changes.



Successful implementation of AI programs requires collaboration between all stakeholder groups. The applications of data science and delivery science are essential. Delivery science focuses on understanding how AI solutions will fit with workflows, how users will interact with the system, and how feedback is monitored to ensure performance. Tools such as stakeholder analyses, risk assessments, and implementation plans are critical.

Interactions between humans and AI systems must also be carefully considered. What will the user-interfaces of AI systems look like and what opportunities do end-users have to provide feedback on them prior to deployment? The intent might be to improve healthcare systems and allow for more time between human clinicians and patients, but could the solution create unintended consequences? Could the solution instead lead to more time spent between humans and computers?

KEY QUESTIONS TO CONSIDER:

1. Who are the stakeholders involved in the oversight of AI at your organization? Are Patients and Families included?
2. Who and through what processes are AI related decisions made in our organization and have these processes and required approvals been communicated and enforced in policy?
3. Who and through what processes will be monitor compliance with decision-making processes and authorities?
4. How will this solution change existing workflows? Who will be impacted and how?
5. How will users interact with the system? Has the solution been designed to integrate seamlessly once deployed?
6. How are clinicians and other frontline staff represented on AI projects?
7. Are all stakeholders impacted by the outcomes of an AI project included in governance and oversight? Have any stakeholders been missed?
8. What consideration is given to the future of work? Is consideration given to how AI may impact job functions and role descriptions? Are the impacted stakeholders engaged in AI work?
9. After implementation, are the same stakeholders involved in monitoring and sustainability of AI solutions that were implemented?
10. What does the feedback process look like to ensure the AI solution is delivering the right results? What is the process for acting and intervening if corrective action is required?
11. If vendors or other external parties are involved in developing a new solution, who owns the solution? Who is liable for the recommendations? Who would bear the cost of any revisions if required?
12. What is the approach for pre-launch trials and evaluation of any AI system? How is the effectiveness of these trials evaluated?

RIGOROUS METHODOLOGY:

Is the right approach being used to solve this problem?

After defining a meaningful question and engaging appropriate stakeholders, the next critical step is defining the approach to solving a problem using AI. This requires understanding the resources available to solve the problem and defining a clear objective for the AI solution to achieve.

As a general example, consider the use case of an AI application built to interpret radiological images. If the AI application detects an abnormal finding, the workflow is to notify a radiologist for their review. How is the effectiveness of the AI application measured? Is it by overall accuracy? Consider the case where abnormal findings are only truly present in 1% of the images. If the AI system declared that no images had abnormal findings, it would only be wrong 1% of the time and would have an "accuracy" of 99%! This example helps demonstrate the need to carefully define the metrics that the AI system will optimize to ensure that the solution provides value when implemented in workflows.

After defining a meaningful question and engaging appropriate stakeholders, the next critical step is defining the approach to solving a problem using AI. This requires understanding the resources available to solve the problem and defining a clear objective for the AI solution to achieve.

The approach to solve a problem using AI will also be dependent on the data available. Is the data used to build and train the system representative of what it will see in real life? Is the data reliable, or does it need pre-processing before it can be used to develop the solution? Are there outliers or unexpected values in the data? How do we even know that a data point is an outlier? How or why do these occur outlier points occur? Do outliers in the data need to be addressed through cleaning or pre-processing before an AI system can ingest it to make decisions? The methods for handling outlier values and for conducting data cleaning and pre-processing have a substantial effect on the overall output and quality of AI models. Will the model have timely access to quality data when it is deployed?

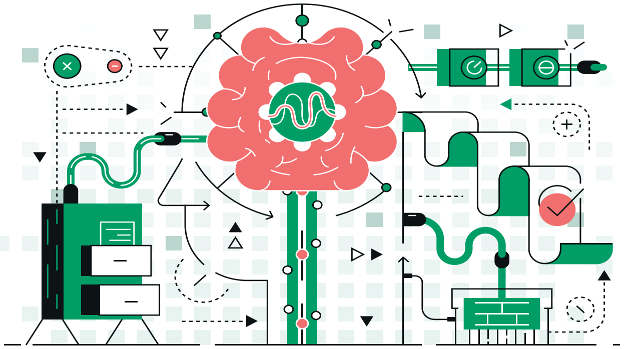
Despite significant advances in recent years, no AI model will produce perfect results. In the example above, the AI system may produce false positive predictions and false negative predictions. Is there one type of error that is worse than others? What performance threshold is tolerable, and how is this threshold determined?

An important consideration is the feedback process whereby AI algorithms and applications are continuously updated as new data becomes available. An example of an approach is Online Machine Learning. This is a method where an algorithm is trained and updated every time a new data point becomes available. This is a more resource-intensive development process but may potentially improve the quality of an AI system.

The amount of data required to build AI applications will vary based on the complexity of the task and intended use of the application. Consider the example of an AI algorithm built to identify lesions in MRI images of brains. This is a complex classification task that may require extremely large amounts of data, perhaps more so than is available at a single organization. In this case, does the organization consider partnerships with other sites to augment datasets?

AI models can be highly complex, and it is not uncommon for these models to have millions of parameters that must be carefully tuned during development. Moreover, programming languages have advanced to the point where it is possible to create functioning AI algorithms using very few lines of code. However, the way AI systems are designed, and the way a model's parameters are chosen will have a large impact on the effectiveness of the model. When creating AI algorithms, programming design choices must be made carefully and deliberately to avoid wasting resources or producing ineffective or even potentially inaccurate solutions.

Any AI solution must be designed carefully around the workflows in which it will be used. Are there unintended consequences of its use? Does a solution that was intended to create more contact time between humans and clients create more contact time between humans and computers?



KEY QUESTIONS TO CONSIDER:

1 What metrics should the application aim to improve? Are these metrics appropriate for the problem being addressed? Do they align tightly with the need for an AI solution?

2 Is the data used to build the AI system representative of data that it will see in real life?

3 No AI model will produce perfect results. What performance threshold is tolerable? How is this determined?

4 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 they provide post-deployment?

5 Before deployment, investigate situations when the model fails. Why does it fail? How are lessons from these failures incorporated into the model to improve performance? What is the feedback mechanism that allows AI algorithms and systems to improve from past errors?

6 After deployment, how does the model continue to learn? What is the ongoing process to collect data and feed it back to the model for continuous improvement?

7 Is there potential for the solution to produce unintended consequences? How are these monitored and measured? Who is responsible for this ongoing monitoring and maintenance?

8 Do solutions that were intended to create more contact time between humans and clients create more contact time between humans and computers?

TRANSPARENCY:

How does AI work and how do we know it's solving the problem?

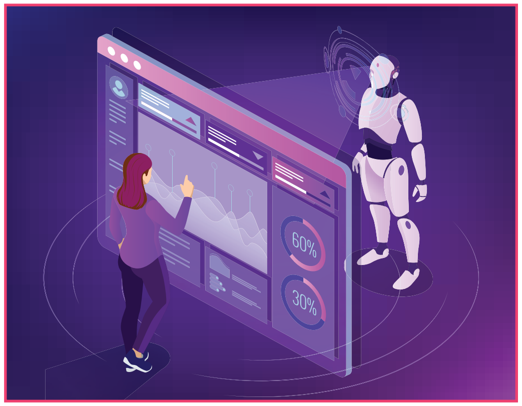
Since the 1960's, humans have built hundreds of thousands of AI models for numerous applications. However, very few of these models are implemented in practice. Two reasons for this are **disuse** and **misuse** of AI:

DISUSE: Building valid and viable models that are not implemented in practice.

WHY DOES THIS OCCUR? Reasons may include that clinicians or end-users may not trust how a model was built, they may not have been involved in the development process, the development work may not address a meaningful problem, and/or not enough consideration is given to practical constraints post-deployment.

MISUSE: Building models that introduce further risks or harm.

WHY DOES THIS OCCUR? Reasons may include that a model was not built using appropriate data, it has inaccuracies, it was built with biases, as described earlier, data that was used is not representative of the actual population, and/or the model development did not consider unintended consequences.



AI systems are probabilistic and are rarely capable of giving precise recommendations. Recommendations from AI systems should be given with a measure of confidence, but who interprets this? Does interacting with AI systems change role responsibilities and job descriptions? Would clinicians be required to explain how an AI model produces a recommendation? The more that clinicians and users feel they understand the overall AI system, the more inclined and better equipped they will be to use it.

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 human clinician, who makes the final decision? These questions and the considerations listed on the right are intended as prompts to help with risk management and planning prior to the start of any AI-based activity at your organization.

KEY QUESTIONS TO CONSIDER:

1. How is the AI system making recommendations or acting? What inputs does the system consider, and how are they weighted?
2. How does the AI system provide reasoning for the recommendations it provides in a way that a diverse user population can understand?
3. What happens if an AI system gives conflicting advice from a healthcare provider? What is the process for resolving differences in recommendations? Is there an escalation process for resolutions?
4. 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?
5. Have clients provided consent to have their data used for development of an AI system? Have they provided consent to have their treatment informed or guided by an AI system?
6. What is the disclosure process if errors occur based on the recommendations of an AI system?
7. During incident reviews and investigations, are the datasets used to train an AI system also disclosed? What if these datasets include personal health information?
8. Who has access to AI datasets when investigating potential failures or adverse events?
9. Is there a process to ensure the legitimacy of the AI solution that is developed? Is the solution built in compliance with relevant legislation, standards and regulations? Is there a credentialing process to ensure compliance?

DATA INTEGRITY, PRIVACY AND SECURITY:

How are we protecting privacy and system integrity?

A unique feature of AI systems is their potential to continuously learn from real-world data to improve their performance after deployment. This is distinct from other models that have been developed and had their parameters frozen before deployment. A critical consideration is therefore how to protect data to ensure its accuracy. Data stewardship is critical to ensure that risks of compromise to data has been minimized.

AI models may also have been built using data from many sources. Where is this data stored, and how is access to this maintained? Who owns the data inventory, and how is it audited to ensure that data assets are protected? Given that AI actions and recommendations taken by AI systems have the potential to impact a vast number of clients, mitigation strategies to prevent vulnerabilities are essential. These vulnerabilities may include cyberattacks, unauthorized access to data repositories, and the integrity and confidentiality of personal data.

> AI systems are highly reliant on the availability of high quality, reliable datasets, both during the development of AI solutions and during in-situ deployment. Even slight perturbations to datasets fed to AI systems can significantly alter predictions or recommendations produced by the system. In a controlled environment, researchers have shown that modifying just one pixel in images ingested into to an AI-based system can greatly alter what the system thinks it is seeing." (Su, Vargas, & Sakurai, 2019).

This example demonstrates the critical need for strong cyber security strategies to protect AI systems from external threats. Threats may include, but are not limited to privacy breaches, unauthorized modification of corruption of data and algorithms, and ransomware attacks.

Additional guidance on risk management for **cyber security** can be found in the following document: HIROC Cyber Risk Management: A Guide for Healthcare Providers and Administrators.

KEY QUESTIONS TO CONSIDER:

1. Is there a master inventory and dictionary of all data assets used by the AI system? How is this maintained and updated?
2. Who and what processes are in place to monitor risk mitigation strategies and any breaches which may occur?
3. How is access to this data inventories maintained? How are these inventories protected?
4. Do additional protocols or mitigation strategies need to be put in place to protect privacy of data used for AI applications in your organization?
5. What measures have been put in place to prevent hardware and software faults that could result in data being compromised?
6. What is the business continuity plan in the event of a service interruption of an AI system?
7. What measures have been put in place to ensure that data is secure? Is there a way to know if data has been corrupted and how?
8. What controls are put in place to manage data during usage, transmission and storage?
9. Is our organization's insurance confirmed in relation to AI initiatives and practices?

Recommendations for putting guiding principles into action

Building on the guidance principles from the preceding section, this section of the guide will introduce minimum recommendations for organizations of all sizes and types to consider when developing or implementing an AI-based solution.

VALUE PROPOSITION: Is AI being used to solve the right problems?

1. Consult with a wide range of stakeholders, including clinicians and other end users to develop meaningful questions to be answered with AI tools and solutions.
2. Consider applications of AI that align with one or more dimensions of a high-quality healthcare system. A high-quality healthcare system is one that is accessible, appropriate, effective, efficient, equitable, integrated, patient-centred, population health-focused and safe (ECFAA, 2010).
3. Create well-defined problem and goal statements based on areas of need and then identify the data requirements to answer the question, as opposed to selecting focus area solely based on available data.
    a. Problem statements should be Measurable, Observable, Specific, Time-Bound
    b. Goal Statements should be: Specific, Measurable, Achievable, Realistic, Time-Bound
4. Obtain feedback from independent stakeholders on the utility and intended applications of potential solutions.
5. Consider tools such as the **Learning Health System**, which provides a framework for creating knowledge from data, translating knowledge into practice, and merging practice into data flow (Institute of Medicine (US), 2007).

GOVERNANCE: Are the right people involved to solve this problem?

1. Consider creating a standalone Steering Committee for AI initiatives at your organization. The mandate of this Steering Committee would include oversight on initiation, planning, execution, monitoring and closeout of an AI projects.
2. Membership of the Steering Committee should include but not be limited to clinicians and other end-users of any AI tools and solutions, as well as Patient and Family Advisors, Ethicists, Risk Managers, and Information Services staff.
3. Carefully evaluate how users will interact with any AI systems and how to ensure seamless integration, including potential changes to processes and functions once a solution has been deployed.
4. Adopt Project Management tools including Stakeholder Charts, Gap Analyses, Implementation Plans, User Acceptance Testing, Testing Plans, and Sustainability Plans for all AI initiatives.
5. Conduct risk assessments to identify potential unintended consequences that may result from the development and implementation of AI tools. Do workflows and processes change as a result? How? What change management strategies are put in place to help mitigate risk in this regard?

RIGOROUS METHODOLOGY: Is the right approach being used to solve this problem?

1. Establish minimum data quality specifications for all AI models or applications. Only datasets that meet these minimum specifications should be used for developing or implementing AI solutions. Data quality specifications should consider accuracy, completeness, consistency, credibility, currentness, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability, recoverability (ISO/IEC, 2008).
2. Carefully select metrics and create a monitoring plan to ensure the AI solution is performing as desired when deployed. A monitoring plan should at a minimum include:
    a. **Outcome metrics:** Measures to assess the objectives of a solution.
    b. **Process metrics:** Measures to know if the system is working as intended.
    c. **Balancing metrics:** Measures to evaluate potential unintended consequences or impacts.
3. Ensure comprehensive use and test cases have been defined, and appropriate testing and acceptance plans have been developed and implemented.
4. Prior to deployment, conduct trials and validation with clinicians, experts and other end-user groups for proof-of-concept testing.
5. Performance of AI systems should be monitored regularly to mitigation potential degradation or bias. Design processes for systematic tracking, reporting, and analysis of errors, near misses, and overrides.

TRANSPARENCY: How does AI work and how do we know it's solving the problem?

1. Develop comprehensive communication plans to ensure all stakeholders are involved and advised throughout the lifecycle of a development project.
2. Consider creating a training plan for stakeholders and end-users outlining how the AI systems function, make decisions, generates insights and recommendations, and how it was developed and how it should be used. Ensure this documentation is easy to access and has been provided to all required people.
3. Develop feedback loops for monitoring performance, usability, and experience of any solution post-deployment. Use this feedback for ongoing monitoring and improvement of the solutions.
4. Develop clear escalation plans, and continuously update this using ongoing collection of feedback on the solution.

PRIVACY AND DATA SECURITY: How are we protecting privacy and system integrity?

1. Create or revise data security plans for any initiative involving AI, which at a minimum includes:
    a. An inventory of data assets
    b. Access permission and controls, including how frequently these should be reviewed
    c. Controls required for transmitting, storing, and accessing data
    d. Data retention and destruction processes
2. Deliver education and communication as required to stakeholders about AI, including threat identification and response plans.
3. Create or revise data sharing and access agreements if required to define access, usage, and ownership privileges to any data assets.
4. Create business continuity and disaster recovery plans for AI tools and solution, including end-to-end infrastructure and resiliency controls.

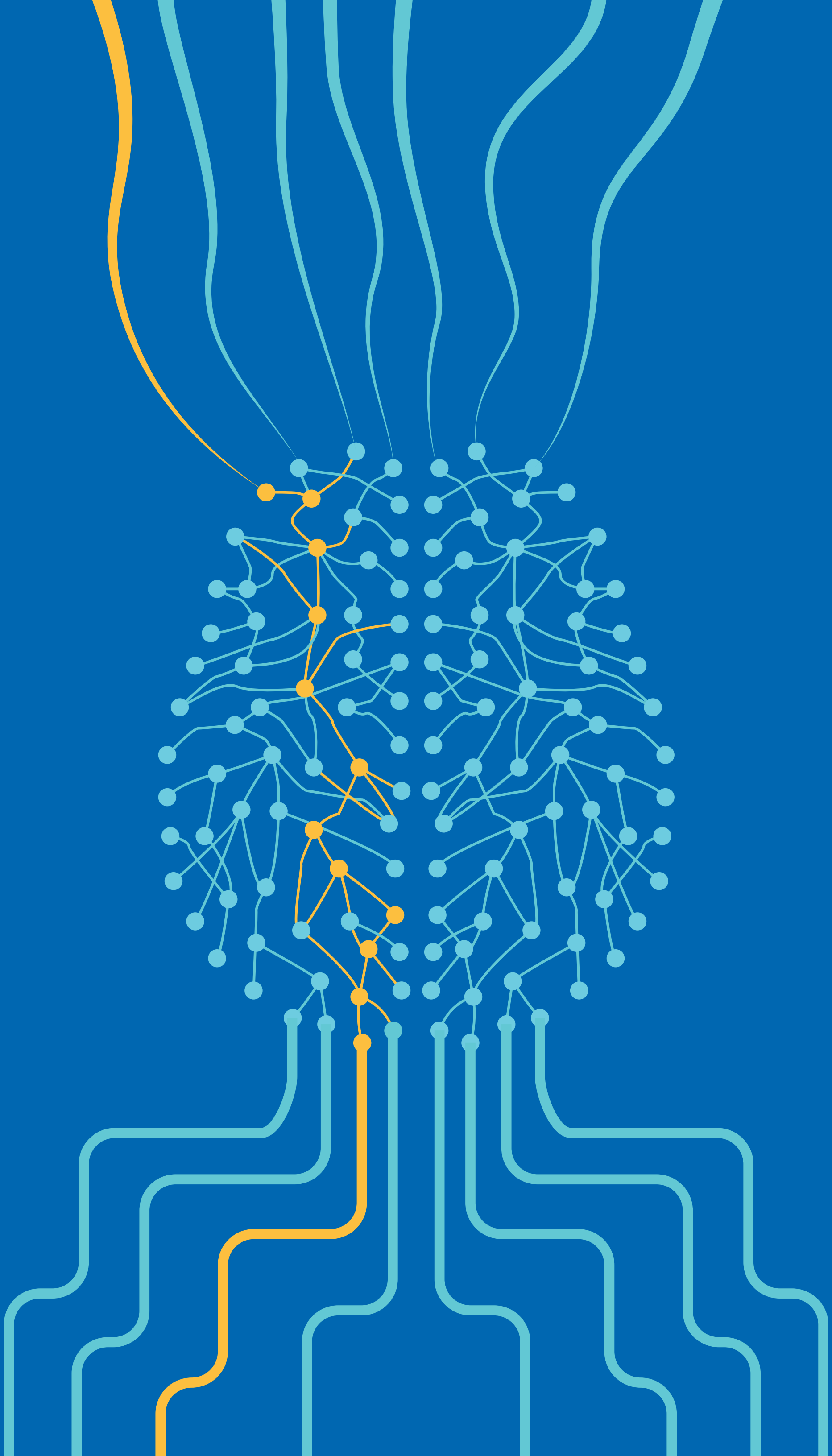
Final Thoughts

Artificial Intelligence has numerous potential applications to improve many aspects of healthcare service and delivery. While the benefits may be large, the adoption and implementation of AI carries new risks that must be carefully identified, mitigated and managed to prevent unintended consequences.

A combination of administrative, logical and technological solutions needs to be employed by healthcare organizations – and even when those steps are employed, your organization must remain vigilant about keeping protective measures current and viable.

This guide is a starting point to help with the management of risks in the adoption and implementation of AI in your organization.

To understand your HIROC insurance policy or learn more about Artificial Intelligence, please contact us at inquiries@hiroc.com.



Glossary of terms

Algorithm
A step-by-step procedure for solving a problem or for completing a task. Algorithms can be applied by data scientists and programmers to help solve complicated analytical problems involving.

Artificial Intelligence
The notion of machines exhibiting and mimicking cognitive functions that are usually associated with humans, such as learning, reasoning, predicting, planning, recognizing, and problem solving. With constantly growing repositories of data, improving sophistication of algorithms and faster computing resources, AI is becoming increasingly integrated into everyday use.

Big Data
This is a term used to describe large, complicated datasets that often include millions of rows and/or thousands of columns. Big datasets are large in volume, often have a variety of data, and are often populated with high velocity. Big Datasets hold the potential for greater statistical power but require more sophisticated storage and analytical tools to process.

Concept Drift
A phenomenon that occurs when the characteristics of a population or process change over time. This may cause problems for AI systems if datasets change in a way that a model was not expecting.

Data Science
A multidisciplinary field of study using methodical and collaborative methods to extract knowledge from datasets. Data science requires knowledge in computer science, mathematics, statistics and engineering to draw conclusions and solve meaningful problems.

Data Warehouse
A centralized repository of data commonly used to generate reports and data analysis.

Decision Support System
A system that supports decision making or business processes within an organization. These systems can span between fully human-controlled and fully-automated.

Machine Learning
The methodical application of algorithms and modeling techniques to analyze data and determine patterns and inferences.

Online Machine Learning
A process where algorithms are trained and continuously updated as new data becomes available.

References

Beyene, A. A., Welemariam, T., Persson, M., & Lavesson, N. (2015, July). Improved concept drift handling in surgery prediction and other applications. *Knowledge and Information Systems*, pp. 177-196.

Canadian Institute for Health Information. (2019). *NACRS Data Elements 2018–2019.* NACRS Data Elements 2018–2019.

Canadian Institute for Health Information. (2019). *National Health Expenditure Trends, 1975 to 2019.* Ottawa: CIHI.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *International Conference on Knowledge Discovery and Data Mining* (pp. 1721-1730). Sydney: Association for Computing Machinery.

Crowe, N. (2019). *Risk-based approaches to artificial intelligence.* London.

Davenport, T., & Kalakota, R. (2019, June). The potential for artificial intelligence in healthcare. *Royal College of Physicians: Future Healthcare Journal*, pp. 94-98.

Denil, M., Shakibi, B., Dinh, L., Ranzato, M., & de Freitas, N. (2013). Predicting Parameters in Deep Learning. *Neural Information Processing Systems Proceedings.* Lake Tahoe: NIPS.

ECFAA. (2010, June 8). Excellent Care for All Act, 2010. Queen's Printer for Ontario.

Ginestra, J. C., Giannini, H. M., Schweickert, W. D., Meadows, L., Lynch, M. J., Pavan, K., . . . Umscheid, C. A. (2019, November). Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock. *Critical Care Medicine*, pp. 1477-1484.

Health Quality Ontario. (2017). *Quality Matters: Realizing Excellent Care for All.* Toronto: Queen's Printer for Ontario.

Institute of Medicine (US). (2007). *Institute of Medicine (US) Roundtable on Evidence-Based Medicine.* Washington: National Academies Press (US). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK53494/ doi: 10.17226/11903

ISO/IEC. (2008). *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model.* ISO/IEC.

Mirsky, Y., & Mahler, T. (2019). CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. *USENIX Security Symposium* (pp. 461-478). Santa Clara: USENIX.

Nicholson-Price, W. (2019). *Risks and remedies for artificial intelligence in health care.* Washington: The Brooking Institution.

O'Malley, K. J., Cook, K. F., Price, M. D., Wildes, K. R., Hurdle, J. F., & Ashton, C. M. (2005, October). Measuring Diagnoses: ICD Code Accuracy. *Health Services Research*, pp. 1620-1639.

Public Health Agency of Canada. (2008). *The Chief Public Health Officer's Report on the State of Public Health in Canada, 2008: Addressing Health Inequalities.* Ottawa: Public Health Agency of Canada.

Russakovsky, O., Deng, J., Su, H., Krause, J., & Satheesh, S. (2015, December). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, pp. 211-252.

Su, J., Vargas, D. V., & Sakurai, K. (2019, October). One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation*, pp. 828-841.

Suresh, H., & Guttag, J. V. (2019). A Framework for Understanding Unintended Consequences of Machine Learning.

United Nations Educational, Scientific and Cultural Organization. (2005). *The Precautionary Principle, World Commission on the Ethicsof Scientific Knowledge and Technology (COMEST).* Paris: United Nations Educational, Scientific and Cultural Organization.

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., . . . Goldenberg, A. (2019, September 19). Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, pp. 1337-1340.

4711 Yonge Street, Suite 1600
Toronto, ON M2N 6K8
416.733.2773
Toll Free: 1.800.465.7357

1200 Rothesay Street
Winnipeg, MB R2G 1T7
204.943.4125
Toll Free: 1.800.442.7751

HIROC.com

HIROC is a trusted healthcare safety advisor, committed to offering a full spectrum of insurance products and support throughout a claim. Combined with sage counsel and risk management solutions, HIROC works with its partners to increase safety.

HIROC

PARTNERING TO CREATE THE SAFEST HEALTHCARE SYSTEM