Enabling Trustworthy and Safe AI in Health Care

Special Report | December 2024

Table of contents

Addressing the current challenges for trustworthy AI in health care

The potential for artificial intelligence (AI) to transform health care is enormous. Yet, the path to realizing these benefits can be complex and challenging. How do we ensure that the AI we trust with our health is trustworthy, safe, and responsible? To unlock AI's potential in health care, critical issues need to be considered, including:

1 Data access

Working with sensitive health data to train and validate AI models requires rigorous data governance, as well as privacy and security measures. Striking the right balance between such measures and accessibility requires careful and nuanced consideration.

2 Bias

If the data used to train AI models is biased, AI can perpetuate and potentially amplify existing biases. Training AI models with biased data, coupled with insufficient bias detection and mitigation in the outputs of that training can result in outcome disparities across diverse populations and inequitable access to AI-enhanced healthcare benefits. Responsible AI development should include checking for bias across the product development lifecycle.

3 AI literacy

Many health organizations lack the necessary expertise and understanding of AI and machine learning to effectively manage and integrate these technologies into health services. To mitigate risks, the core elements of trustworthy and safe AI need to be understood by the teams responsible for managing and utilizing AI-enabled solutions in health environments.

4 Transparency

The black box nature of some AI algorithms can make it difficult for health care providers to understand how decisions or predictions are made. A lack of transparency diminishes trust in AI systems and can be an impediment to the validation and verification of AI outputs in clinical settings. AI models performance should be transparent and explainable to end users.

5 Accountability

When deploying health AI solutions in clinical contexts, ownership, accountability, and governance need to be well defined and understood by all individuals involved to ensure that responsibility for ethical application of technologies, legal liability, and professional accountabilities are clear. Accountability and risk management protocols should be preemptively established to address potential AI failures or harms before they occur.

6 Adaptive regulation

AI advancement is outpacing the existing regulatory landscape, resulting in gaps in governance. This can lead to difficulties in ensuring that AI technologies meet the highest standards of safety. Conversely, overly stringent AI regulation could stifle innovation and translate to delays in adoption of beneficial technologies. An adaptive AI regulatory environment can encourage innovation and evolve as needed with technology and its application to different sectors.

VECTOR INSTITUTE | INSTITUT VECTEUR

Building from the Vector Institute's Trust and Safety Principles

The Vector Institute's mandate is to drive excellence and leadership in Canada's knowledge, creation, and use of AI to foster economic growth and to improve the lives of Canadians. Vector has shown leadership in turning its advanced health AI research into reliable solutions that address health care system challenges and work towards improving health outcomes for all Canadians. To ensure that organizations adopt safe approaches for developing and deploying AI in their workflows, Vector has successfully implemented a number of initiatives that are anchored in its AI Trust and Safety principles. In this special report, we refer to "responsible AI" as an approach to developing, deploying, and maintaining AI systems in a safe, trustworthy, and ethical way.

Released in 2023, Vector's AI Trust and Safety Principles build upon the ethical AI approach developed by the Organization of Economic Co-operation and Development (OECD).

While the principles are a useful starting point, they are only one part of the solution. To promote responsible AI deployment, these considerations must be translated into concrete actions. Organizations should adopt similar principles and look to frameworks for steps to creating and implementing responsible AI. Since the establishment of its AI Trust and Safety principles, Vector has developed two key resources:

1. Responsible AI Product Development framework to guide developers, designers, engineers, and researchers on how to build AI responsibly throughout the product development lifecycle; and

2. Health AI Implementation Toolkit with implementation journeys, checklists, and safe AI considerations to guide the deployment of AI-enabled solutions into clinical practice or administrative functions in health services.

VECTOR INSTITUTE | INSTITUT VECTEUR

Industry spotlight:
EY's responsible AI framework

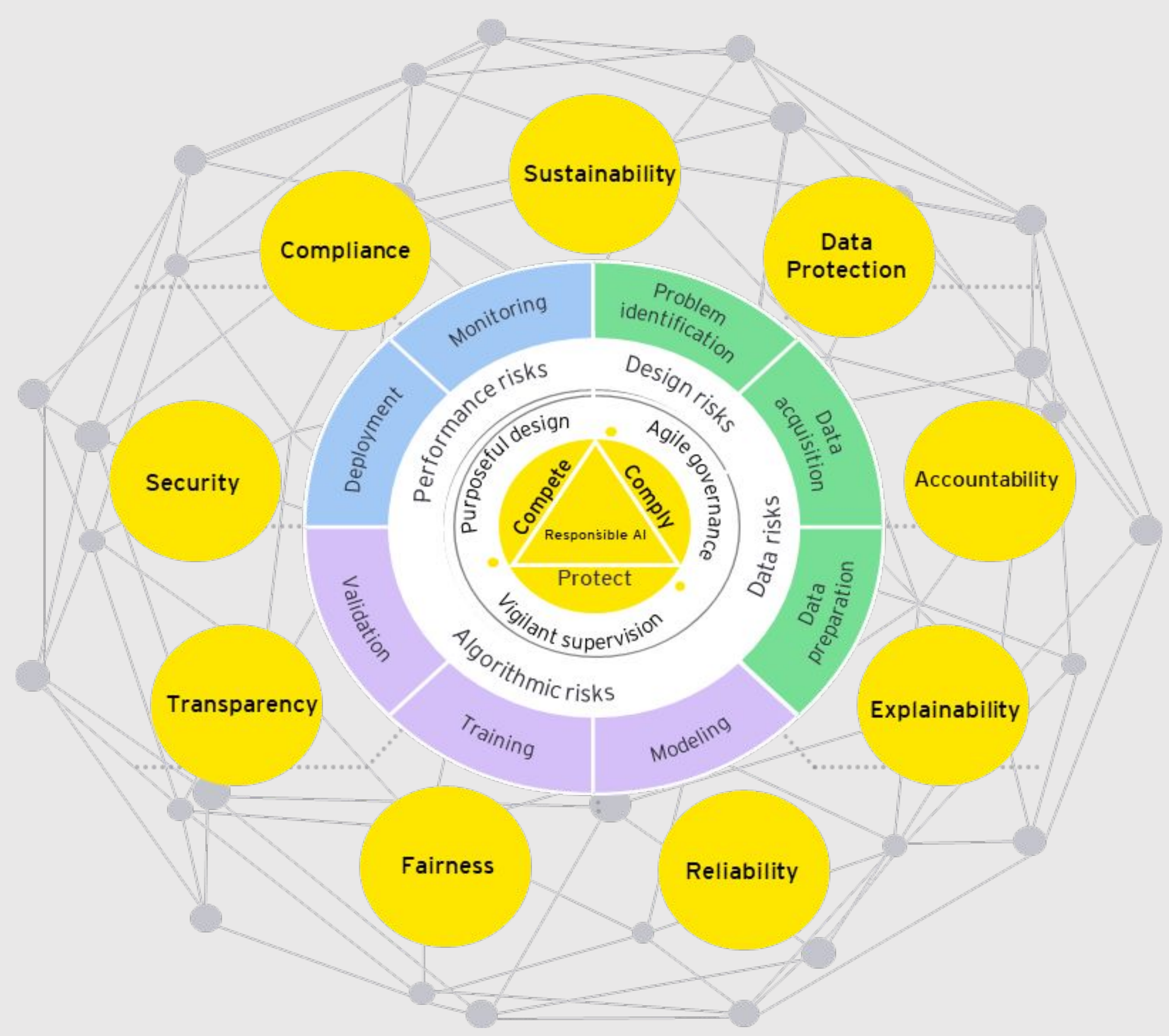

Figure 1: EY's responsible AI framework based on the key foundational elements of purposeful design, agile governance and vigilant supervision.

Vector Gold sponsor EY Canada's publication, "Six ways to make more of AI in Canadian healthcare," identifies six critical areas for responsible AI integration, offering a comprehensive framework to guide health care organizations in leveraging AI for improved health outcomes. As AI advancements and public accessibility surge, EY's insights pave the way for a future where AI-driven solutions redefine patient care and operational efficiency across Canadian health care.

On May 13, 2024, Vector and EY Canada co-hosted an event called, "Principles to Practice: Enabling Responsible AI in Healthcare." The event brought together over 80 senior health care leaders from across Ontario to learn about implementing AI and to hear expert opinions on the pace of safe AI deployment in health.

CyclOps for clinical model monitoring and evaluation

Vector is working to make the post-deployment stage of product development easier and safer for all users. The "Principles to Practice: Enabling Responsible AI in Healthcare" event highlighted CyclOps, an open source tool to help scientists, engineers, and clinicians evaluate and monitor machine learning (ML) models in clinical settings. Vector researchers who have used the tool on their deployed models presented their project journeys during the clinical use case presentation segment of the event.

Clinical use case spotlights

Delirium prediction

Predicting hospital-acquired delirium with AI-based tools

Presented by Dr. Amol Verma (Unity Health Toronto, Co-founder of GEMINI; Faculty Affiliate, Vector Institute) and Dr. Fahad Razak (Unity Health Toronto, Co-founder of GEMINI; Faculty Affiliate, Vector Institute)

Delirium is an acute confusional state that has sometimes been called "brain failure". Identifying cases of delirium is a challenge in hospitals across Ontario. Delirium is increasingly recognized as a growing, yet preventable, cause of harm, with approximately one in three patients admitted to a medical ward developing delirium.

While delirious, individuals can become unable to recognize their loved ones, unable to eat or sleep properly, and can get worse over time. Delirium is highly distressing for patients, caregivers, and providers.

In addition to the emotional burden, delirium also negatively impacts the health care system.

On average, patients who are delirious stay eight days longer in hospital, have twice the mortality rate, and cost approximately $11,000 more per hospitalization than the average patient.

Across Canada each year, that works out to approximately 500,000 people and $5 billion dollars.

Delirium can be prevented in 20 to 40 percent of cases with timely intervention. But as Dr. Razak noted during their presentation, prevention is challenging. "It involves what at first seems like very simple interventions like eating, moving, helping people sleep better, keeping them hydrated, helping them with sensory aids, helping them with orientation," he said. "But the problem fundamentally is that these are human interventions and human resources situations across many hospitals [are] very, very constrained." Moreover, overburdened staff cannot always know which patients are at risk of developing delirium.

For the Unity Health team, this challenge presented an opportunity to apply ML. Their goal was twofold: to identify patients currently experiencing delirium and to predict those at risk of developing it.

This approach would allow staff to target interventions and resources towards the highest-risk patients. Foundational work included trying to better capture delirium in hospital records. At base, administrative diagnostic codes capture only about 25% of delirium cases. This was the first application of ML. "To take all the data that exists in a hospital's electronic medical health record and use it to classify when a patient did, and did not have delirium, (it) turns out we can, and we can do it very reliably," said Dr. Verma. The team collaborated with engineers at the University of Toronto to develop an ML tool that demonstrated high performance in measuring delirium rates, retrospectively.

Prior to the application of AI, the best method of identifying delirium for quality measurement or research has been: "trained clinicians go through the medical records and in detail read all the clinical notes," said Verma. "And then to determine whether or not a patient had delirium, they look for symptoms like confusion, encephalopathy." He describes this process as labour intensive and time consuming, ranging from 30 minutes to two hours per chart review, which is not a scalable way to identify when patients have delirium.

Instead, the team trained assistants that reviewed and labelled thousands of these records, in order to train an ML model to recognize and classify delirium in an automated way. This demonstrated that an ML model could measure delirium rates accurately within a hospital, and more importantly, could classify which patients did and did not have delirium.

The team is now focused on developing a real-time ML prediction tool. This involves using all the information available in the electronic medical record to create a dynamic tool that can, at the time when a patient is admitted to the hospital, integrate all of these factors and predict who is at risk of developing delirium. A prediction score would classify patients as high, medium, or low risk, enabling earlier intervention and improving patient outcomes.

Razak and Verma highlighted the value-add from Vector's CyclOps framework, from dataset preparation to model evaluation and monitoring, to improve prediction rates. The team is working towards prospective validation of AI-enabled delirium prevention across Toronto Academic Health Science Network sites, and has plans to pilot deployment at three to five hospitals in 2024.

The team is also developing a quality improvement strategy and intervention funnel for patients identified as being at the highest risk for delirium.

In closing, Verma urged the audience to continue thinking about the clinical opportunities of AI and ML technologies and how their use can reduce the burden of preventable harm, and improve quality of life for patients and the clinicians caring for them.

Stroke CT monitoring

Monitoring third-party stroke CT prediction model

Presented by Dr. Benjamin Fine (Trillium Health Partners, Clinician Scientist and Radiologist; Faculty Affiliate, Vector Institute)

In-house development and deployment of health AI models require specific expertise and resourcing that may not always be available. As such, many AI-enabled solutions implemented in health care environments today are delivered through commercial vendors. This raises the questions: how do you ensure procured models will work well locally? How many such models can continue to perform well outside of the data on which they were trained?

According to Dr. Fine, the answer is not that many. Currently, there are no regulatory requirements in place for deployed and approved models to have continuous performance monitoring.

“Once you change the underlying data, you have basically no idea how exactly the model will perform,” said Fine. He emphasized that health care AI is high stakes AI, where false predictive outputs or decisions can have drastic consequences.

Fine’s team at Trillium Health Partners (THP) is using a commercially-developed tool to look at the clinical problem of acute stroke. The treatment and management of stroke patients is a highly time-sensitive task — even small reductions in time to treatment can mean significantly better outcomes for patients.

As Fine put it, “time is brain”, and in order to identify who is most eligible for life-saving treatment (such as the insertion of a catheter into the brain to pull out a clot), specialized imaging is needed. This imaging process takes time and the human resources to go through thousands of images and determine which patients are candidates for treatment.

Fine described his team's approach as, "Day 0: what you need to do before you are deploying a tool; day 1: when you turn the tool on and it's working in the background; and day 2: when you are actually using the tool with humans and would like to know how the tool is performing." Governing these technologies from a people and process perspective is important, but monitoring the models across the full lifecycle is imperative.

Fine's team collaborated with Vector and leveraged the CyclOps framework to evaluate and monitor the performance of the tool. The Vector and THP teams worked to develop a type of dashboard that displayed performance metrics of algorithms that were deployed. While the design of the interface included things like overall performance versus baseline and literature, performance over time, and data drift, Fine suggested that the most important part of such an interface is the manner in which these risks are communicated to the people who are ultimately going to be using these tools. Similar to a drug insert, the teams integrated a model fact card into this interface to give users an idea of where a model could work, where it could not, as well as any risks.

The next steps for the THP team include applying the CyclOps framework to other deployed AI tools, and refining automations for a natural language processing tool to extract ground truth from radiology reports.

CORAL

Enabling automated detection of pneumothorax with AI

Presented by Dr. Michael Brudno (Chief Data Scientist, University Health Network; Faculty Member, Vector Institute) and Dr. Chris McIntosh (Scientist, University Health Network; Faculty Affiliate , Vector Institute)

Pneumothorax is a life-threatening event resulting in a collapsed lung. It often requires the insertion of a chest tube to resolve. While this can seem quite obvious in a diagram, it is not as easy to decipher from an x-ray scan — especially for non-clinicians. "It can be hard to see," said Dr. McIntosh. "You need a well trained radiologist, and that is all well and good if you get to a hospital downtown at say 2 o'clock in the afternoon. There's a large roster of radiologists to read your scan. But if you unfortunately come in at 2 in the morning, there can be a much smaller roster available."

McIntosh noted that what often happens in these cases is that seemingly less urgent scans are not read until the next shift comes in. This prompted an idea for the University Health Network (UHN) Data Aggregation, Translation and Architecture (DATA) team to create an automated prioritization tool. "What if we can build a tool to alert the existing radiologists — the on-call ones — at night or at 2 in the morning to say there's something urgent that could use an intervention, rather than waiting until 8 a.m. to see a particular scan?"

To realize their idea, the team faced a few challenges. The first challenge highlighted by McIntosh was that "AI doesn't generalize well a lot of the time." He cited a systematic review from 2022 that looked at the internal validation of published papers and then compared it to the accuracy when externally validated with other hospitals. On average, "they dropped around 20%," he said. For clinical tools, not requiring continuous monitoring post-deployment is dangerous. "We have a particular view of hospital data where, let's suppose we have x-rays from the ER and a whole bunch of x-rays from the cardiology department. What we tend to assume is that there is a consistent split between those two things in terms of positive and negative data — that is, that the data itself is fair and unbiased." The reality is that the majority of positive cases may come from the ER and the majority of negative cases may come from cardiology. And the AI model ends up learning how to detect where the patient came from (whether they went through ER or cardiology), less so than the disease. One of the ways to overcome this is using data from many centres. This is what the UHN team did, although they started off without any internal data.

The team used four public datasets from four hospitals — about 181,000 scans in total — to build their tool. McIntosh described this approach as, "training externally to deploy internally." The model was validated at UHN on about 2,200 scans and found to be about 90% accurate in the testing phase of the external data, and 87% accurate when deployed internally at UHN.

However, the team discovered that in many positive cases of pneumothorax, there is a presence of a chest tube which seemed to be what the AI was identifying. A shortcut now existed not only with external data, but also internal data showing positive cases post-pneumothorax with the presence of a chest tube. There is limited to no value for clinicians to have a model that identifies patients who had pneumothorax but were already treated or cured.

Addressing this challenge, Dr. Brudno shared the team's solution, demonstrating that “two AIs are better than one.” The team built two AI models, one to predict the presence of pneumothorax, and one to predict the presence of chest tubes. While one might expect information about pneumothorax treatment to be readily available in medical records, Brudno explained that these records are not easily accessible and are often delayed.

The team put the two models together in their pipeline to first detect the presence of a chest tube.

If a chest tube was detected, then the scan was marked as negative and not prioritized. If no chest tube was detected, the model would then check for pneumothorax and flag as positive or negative. The priority tags would then be visible to the radiologists through UHN’s radiology dashboard, called CORAL, which is a key component of this project.

The final step in deploying the model was monitoring. Monitoring and evaluation are critical for models to continue supporting health practitioners with workflow, and to make safe and accurate triaging decisions. Vector’s CyclOps framework was deployed at UHN to monitor model performance over time and generate weekly monitoring reports.

Panel on pace and principles: Balancing safety and speed of AI development and deployment in health

The debate on technology acceleration versus deceleration, particularly as it relates to AI, is becoming increasingly relevant. In highly complex and sensitive sectors such as health, there is pressure to accelerate advancement where there is potential for benefit. Does slower development conflict with our ethical obligations to improve health outcomes and quality of health care services? How do you control risk and responsibility when things go wrong as a result of hasty deployment?

Moderated by Safia Rahemtulla, Partner and Public Sector Risk Leader at EY, a panel of experts shared their thoughts on the pace of AI development and deployment in health during the Principles to Practice event.

The panel included:
- *Cathy Cobey, Global Responsible AI Co-Lead, EY*
- *Dr. Devin Singh, Faculty Affiliate, Vector Institute; Paediatric Emergency Physician, The Hospital for Sick Children; Co-founder & CEO, HeroAI*
- *Dr. Jennifer Gibson, Director, Joint Centre for Bioethics, University of Toronto*

Defining responsible AI and the intersection of speed in development and deployment

Responsible AI takes a human-centred and outcomes-focused approach. Dr. Gibson highlighted the importance of responsibility as an ethicist and that responsibility has less to do with the technology itself and more to do with the human beings using these solutions. She noted that Vector exemplifies this commitment by collaborating with others to achieve meaningful, people-centred outcomes in a transparent way – acknowledging the potential for errors and implementing mechanisms to respond to them as they progress.

Responsible AI understands urgency. Dr. Singh looked at the problem from the perspective of an emergency physician. “There is a responsibility to shepherd this technology urgently,” he said, noting that this does not mean without due diligence or incredible thoughtfulness. Instead, he is “seeing and feeling the challenges on the ground everyday. With wait times increasing and larger patient volumes, we're quite literally seeing the harms."

Responsible AI addresses a need and provides real value to users. Cobey spoke about the need to broaden the definition of responsibility to include not only fairness and explainability but also security, data protection, and sustainability. She cautioned, however, that we need to constantly balance "our roles as custodians of health data and innovators of our health services." She suggested bringing this perspective to the use case selection process and to look at the value of those use cases when deciding how to proceed. When different types of clinical, technological, and research investments are made, how is value being measured? How does one measure effectiveness, or benefits to overall health and wellness or patient value? Cobey also noted that we should be clear on the benefits — investment into research should not be driven by solely interest in a specific area, but also the benefit that it can provide to the larger community.

Implementing AI in the clinical domain — other obstacles to deploying AI

Governance and leadership of AI projects, as well as strong change management practices, remain challenges in deploying AI within a clinical context. As an AI researcher and industry co-founder, Singh described the challenges in obtaining approval of an AI project at different institutions. "There lacks a clear runway or process. Who is the right person to say 'yes' to this project?" One of the obstacles that he overcame through his work at SickKids was establishing appropriate governance structures. "How do we rethink an AI REB? How do we think about privacy, security, and cybersecurity infrastructure? How do we create policy around improving these projects so we can go quickly from project idea to data access to then at least proof of concept around a model that may have deployment benefits?" Singh also reiterated the importance of change management. "You can build these incredible models, but the change management is critical to actually getting the uptake," he said. It starts with saying to the end user, "this is the platform we can provide. What do you want the workflow to be?"

Singh cites an example from his institution, SickKids, where nurses from the mental health unit completely redesigned a workflow for redirecting patient alerts within the emergency room. This change reduced the length of stay for these patients by about two hours. Singh reiterated that, “this was not a workflow that an AI leader in the hospital would have come up with. This is a workflow that was empowered by those on the ground.”

Regulatory frameworks and governance models for a balance between responsible AI in health care and agile innovation

Embracing a learning mindset and setting different priorities for regulation and governance can be key to enabling responsible AI and agile innovation. Gibson reinforced that we are at a stage of social innovation around governance. There are institutions with great research ethics governance and clinical governance, but as learning health systems, we do not have the governance to support AI. There are examples of institutions that are innovating while establishing pathways for governance among their own internal teams or, like Vector, releasing tools that support the process. People are becoming increasingly aware that it’s not enough to simply deploy an AI solution or implement one from a vendor, expecting it to work indefinitely—this approach often falls short. While governments are coming out with different regulatory guidelines, there is a need for implementation insight on how to translate guidelines into context. However, it can be problematic to adopt one set of guidelines for all environments; challenges exist within unique contexts, irrespective of technology. There is a lot of learning happening in institutions and a willingness to share those lessons, but to have a population health impact, we probably need to set different priorities.

Insights from other industries on responsible deployment of transformative technologies

There is a lot to learn from industries that have established collaborative and cost-effective data sharing mechanisms. Cobey urged healthcare organizations to take a step back and not overcomplicate data privacy issues. She suggested that surveys have shown that citizens are more comfortable with their data being used in support of positive health outcomes for themselves and their community than healthcare organizations are currently leveraging. There are ways to think differently about capturing information from data. "Think about the minimum amount of data you would need and have a data exchange system — a central body that can set the policies, standards, and be the interaction hub. It could also act as a central data repository where, as the patient, you can access the data, review it for completeness and quality, and manage the permissions to give it out." Cobey envisioned health care being the gold standard for achieving improved data access and effective data sharing. Governance is costly, so there is a need to ensure that it is prioritized for applications with the most risk.

Ensuring that AI systems are transparent, explainable and accountable without slowing down the innovation process

The panelists had three key takeaways for health care organizations to ensure that AI systems remain transparent, explainable, and accountable without slowing down the innovation process:

1. **Be realistic about speed and capacity.** Gibson urged the audience to prioritize leadership and health care system capacity, recognizing that we often have to go slow to go fast.

2. **Leverage basic technology for better governance of AI.** Currently, Health Canada does not allow approval for any adaptive AI, "but all models are going to deteriorate eventually and the process is a bit backwards," said Singh. We can look at solving for more rigorous governance with AI automating this process. If the right investments are made, speed and quality are achievable outcomes.

3. **Be comfortable with failing fast.** Models need to be moved out of testing labs and used against production data sooner. It is also important to avoid a one-stop model for everything. Cobey emphasized that context is key and models should be limited to what they do well.

Vector's solutions

CyclOps

The clinical use cases and panel insights shared during the Principles to Practice event highlighted that there are many challenges to ensuring the strong and reliable performance of ML models post-deployment. Establishing robust and consistent monitoring and evaluation practices for deployed clinical solutions is key to responsible AI in health care, and to maintaining synergistic relationships between clinicians and technology.

Vector continues to iterate on its machine learning operations (MLOps) tool, CyclOps, ensuring new versions and developments are reflective of evolving clinical monitoring environments. Currently, teams at Vector are building a locally deployable application for monitoring metrics like model health in the form of a user-friendly dashboard. This new addition will allow users to be alerted when there is a performance drop on any deployed model.

Federated Learning for Healthcare

Deep learning models require large volumes of data to perform well. In a health environment, where data can be fragmented and barriers to data access may exist, federated learning (FL) presents a unique opportunity. Using FL enables distributed clients to collaboratively train a model without directly sharing their private datasets. FL can help ensure that sensitive patient information is protected, thus aligning with privacy regulations like General Data Protection Regulation (GDPR) and Personal Health Information Protection Act (PHIPA). This decentralized approach also reduces the risk of large-scale data breaches, as the attack surface is limited to individual data sources, making the overall system more secure.

In addition to privacy benefits, FL has also been shown to outperform centrally trained models.[1] Training models across diverse datasets from multiple institutes can aid in mitigating bias and can improve generalizability.

Vector has developed an open-source library, FL4Health, taking FL research to application with straightforward and composable modules specifically designed for health care. FL4Health makes FL research and deployment easier, more robust, and reproducible. FL could be a critical enabler to delivering more accurate, equitable, and secure AI-driven insights.

1. Tavakoli, Fatemeh, D. B. Emerson, Sana Ayromlou, John Jewell, Amrit Krishnan, Yuchong Zhang, Amol Verma, and Fahad Razak. "A Comprehensive View of Personalized Federated Learning on Heterogeneous Clinical Datasets." arXiv.org, July 4, 2024. https://arxiv.org/abs/2309.16825v3.

Monitoring and evaluation of large language models (LLMs)

Vector's Unbias toolkit is a generative AI tool developed by Vector to identify and neutralize biases in text datasets. The toolkit ensures underlying linguistic data is free from biases, leading to fairer outcomes. This customizable toolkit can be tailored for specific health data needs. It is capable of performing multi-modal analyses, allowing bias detection in both text and images to highlight any disparities that require immediate attention.

The Unbias toolkit is also geared to identify and correct misinformation in health-related content. This ensures that the information guiding public health decisions and individual choices is factual and trustworthy.

As LLMs become mainstream in health care, there is a need to develop a gold standard approach for monitoring and evaluating these systems effectively to mitigate bias and improve transparency. This is an active area of work at Vector.

Where do we need to go?

Our vision for responsible AI

To strike the right balance between responsible AI approaches and faster adoption and deployment of solutions, there are three main target areas to consider: enabling adaptive ML regulation, facilitating AI deployment at scale, and developing structure for AI accountability.

Adaptive ML regulation

Developers, researchers, and adopters of AI-enabled solutions in health care need regulatory frameworks that are robust, adaptive, and clear to follow. The U.S. Food and Drug Administration (FDA) and Health Canada have introduced the concept of a "Predetermined Change Control Plan" (PCCP)[2] for AI/ML-enabled device software functions, which outlines how a manufacturer intends to modify their ML model after initial approval. Vector supports the implementation of PCCPs for licensing as ML components of a medical device require model retraining and maintenance. Manufacturers must describe the need for changes to the model after its initial deployment. This description should be linked to the risk associated with potential failures or sub-par performance, ensuring that updates are both necessary and appropriately mitigated.

The FDA currently allows updates to AI models in use, but only within pre-approved guidelines. Changes outside these guidelines likely require a new approval process. While these guidelines are meant to be flexible, they can create challenges in keeping AI models up-to-date. For instance, if urgent safety changes fall outside the approved scope, waiting for a new approval could delay fixes, potentially affecting patient safety and how well the device works. Additionally, needing new approvals for each change could slow down the use of the latest AI advancements in existing devices, which might put companies at a disadvantage globally. Guidelines that cover a wider range of potential changes are encouraged, but this raises the question: Is there a more efficient way to ensure these changes are safe and effective?

The pre-deployment phase should include a silent trial to see how the AI model performs in practice, beyond theoretical testing. Many AI models are used based on published studies and technical reports, but often there is no rule saying that they must show how well they work in the specific setting in which they will be used. Models trained on one set of data might work differently in a new environment with different types of data. So, including a silent test could give valuable insights into how the model actually performs in the real world.

2. Canada, Health. "Government of Canada." Guiding principles: Predetermined change control plans for machine learning-enabled medical devices - Canada.ca, July 29, 2024.

Another crucial aspect is creating a system to check how well an AI model is actually working after it is put into use. This system, called a ground truth pipeline, often doesn't exist for models that have not been developed in-house. Such a system allows quick access to accurate performance benchmarks, helping to track and adjust the model over time. Without this pipeline, it might be hard to know how well the model is really working, which could make it less reliable.

To check how well AI models work after they are put into use, and to show why improvements are needed, it could help to use technology alongside current guidelines. For example, reports that show how the model is performing in real-time could be implemented.

It would also be useful to have a shared system that shows how models work in different places, like big city hospitals versus small rural ones. This would help users understand how the model might work differently in various settings and guide future updates. As such systems are implemented, the rules for overseeing these AI medical tools need to change to fit their unique features, while still keeping necessary safeguards in place.

Deploying at scale

With Canada's publicly-funded health system, and access to the right talent and infrastructure, the country has the ingredients for successful AI integration in health. Support and incentives for AI leaders in the public and private sectors are needed to make meaningful advancements in patient care and system efficiency. However, building out a data science team at scale for every organization is neither sustainable nor financially feasible. Instead, what is needed are federal and provincial government plans to support centres of excellence that can best use and distribute resources.

Strategic expansion of high performance computing infrastructure required for ML is essential to avoid resource constraints and to empower Canadian AI leaders to remain at the forefront of innovation. Leveraging economies of scale will promote the economic and social benefits of AI commercialization. "Test and tries" should be conducted in a controlled environment with experts, researchers, clinical champions, and end-users engaged from the start. Successful proof-of-concepts can then be scaled to other centres, ensuring sustainable and efficient deployment of AI at scale across Canada.

Readying your organization for AI adoption

So, what does an organization need to be ready to adopt AI? Whether an organization chooses to build or buy, local data scientists and ML experts are integral for effective validation, change control, intervention, and monitoring. It is crucial to account for the ongoing support and maintenance required for both the AI model and the computing infrastructure running it. Organizations should also have a basic level of computing capacity available either on-premises or in commercial cloud environments. Developing and running AI models often involves processing large amounts of data, so infrastructure must be sufficient to handle the necessary data volume and velocity. Finally, it is important to consider future needs and to ensure that the selected computing solution can scale as AI usage grows.

AI accountability

AI accountability is the key to the successful and responsible integration and operation of AI systems. This should be approached through three main pillars: change management, risk management, and AI literacy.

Employing strong change management principles from the beginning of an initiative is essential for success. This means that everyone who interacts with or is affected by an AI system must be included in the product development lifecycle. Establishing robust feedback mechanisms so that the AI system evolves in alignment with user needs and expectations is key. Continuous feedback helps identify issues early, provides improvement opportunities, and reduces the likelihood of failed deployment. These mechanisms should also factor in real-time monitoring of an AI system's performance. By focusing on comprehensive change management, organizations can ensure smoother transitions and higher adoption rates of new AI technologies.

Risk management is fundamental to AI accountability, requiring well-defined control measures to address potential risks. A structured approach for managing risks associated with AI deployment is crucial. This includes having clear protocols for different scenarios such as low, moderate, and worst-case outcomes of AI performance failures. Responsibilities for maintenance, change management, and decision-making must be decided before deployment so that organizations can respond swiftly if issues arise.

Such preparedness not only mitigates the impact of potential errors, but also reinforces trust in AI systems. It allows for faster implementation and iteration, as risks are managed proactively, enabling a more agile development process.

AI and data literacy within an organization are critical to upholding strong change and risk management. Education and training programs are necessary to equip stakeholders with the knowledge and skills required to interact effectively with AI systems. Empowering stakeholders to provide input and share their perspectives fosters a culture of inclusivity and collaboration. When users understand how AI works and the data upon which it relies, they are better positioned to contribute meaningfully to its development and evaluate its performance accurately. This shared understanding accelerates the identification of areas in need of improvement and supports rapid iterations. An AI literate team can quickly adapt to changes and implement new solutions, driving faster and more efficient AI integration.

Achieving a balance between the need for responsible AI and the demand for swift deployment requires a multifaceted approach. Adaptive regulation, scaled deployment, and robust accountability frameworks are essential for guiding AI innovations responsibly. Implementing PCCPs can ensure ongoing safety while accommodating necessary updates, yet there must be room for flexibility to avoid hindering advancements. To use AI widely and effectively, organizations need to manage their resources carefully and work together with other expert groups to make the best use of their combined skills and tools.

Finally, preparing organizations for AI adoption involves not just the integration of technical resources, but also cultivating a culture of change and risk management through AI literacy. By fostering a comprehensive understanding and proactive management of AI systems, stakeholders can drive successful and responsible AI integration. In summary, the approach must be dynamic, inclusive, and forward-thinking, aligning regulatory practices with practical deployment strategies and accountability mechanisms to harness AI's full potential while safeguarding its impact.

Thank you to our contributors: Roxana Sultan, Ryan MacDonald, Janice Da Silva, Anushka Rajaram, as well as Dr. Amol Verma, Dr. Fahad Razak, Dr. Benjamin Fine, Dr. Michael Brudno, Dr. Chris McIntosh, Cathy Cobey, Dr. Devin Singh, and Dr. Jennifer Gibson.

Vector would like to express gratitude to all of the attendees and participants for sharing their knowledge and time, and to its sponsors and funders.

About Vector Institute

Launched in 2017, the Vector Institute works with industry, institutions, startups, and governments to build AI talent and drive research excellence in AI to develop and sustain AI-based innovation to foster economic growth and improve the lives of Canadians. Vector aims to advance AI research, increase adoption in industry and health through programs for talent, commercialization, and application, and lead Canada towards the responsible use of AI. Programs for industry, led by top AI practitioners, offer foundations for applications in products and processes, company-specific guidance, training for professionals, and connections to workforce-ready talent. Vector is funded by the Province of Ontario, the Government of Canada through the Pan-Canadian AI Strategy, and leading industry sponsors from across multiple sectors of Canadian Industry. For further information or media enquiries, please contact: media@vectorinstitute.ai

VECTOR INSTITUTE | INSTITUT VECTEUR