April 2023

Financial Industry Forum on Artificial Intelligence:

A Canadian Perspective on Responsible AI

OSFI
BSIF

GRI | GLOBAL RISK INSTITUTE

Contents

04 Foreword

05 Executive Summary

08 Introduction

10 Explainability
- 11 Explainability Dimensions
- 13 Explainable AI
- 15 Disclosure
- 17 Explainability and Trust

18 Data
- 20 Data Characteristics and Associated Challenges
- 22 Data Governance
- 24 Third-Party Data
- 25 Alignment of Data and Business Strategies

26 Governance
- 27 An AI Governance Framework
- 29 Evolving Existing Governance Frameworks for AI
- 32 Skills, Culture, and other Challenges
- 34 The Way Forward

36 Ethics
- 37 Legal, Policy and Regulatory Implications
- 38 Challenges and Considerations
- 39 Business Goals, Data Bias and Unfairness
- 40 Fairness
- 43 Privacy and Right to Recourse
- 45 Operationalizing AI Ethics and Organizational Structures

46 Regulations
- 47 State of AI Regulations and Policies across Jurisdictions
- 48 Characteristics of Successful Regulations
- 49 Voice of the Regulators

54 Conclusions

56 Appendix
- 57 Acknowledgements
- 58 Keynote Speakers
- 59 Speakers
- 60 Participants

62 References

Disclaimer

The content of this report reflects views and insights from FIFAI speakers and participants.

The content of this report should not be interpreted as guidance from the Office of the Superintendent of Financial Institutions (OSFI) or any other regulatory authorities, currently or in the future.

Foreword

The rapid advancement, proliferation, and transformative nature of the artificial intelligence (AI) technology has accentuated the need to consider ethical, legal, financial and social implications of its development and deployment. This is where well-designed risk management practices come in, comprised of robust testing and evaluation frameworks, implementation of clear and transparent decision-making processes, and creation of mechanisms for accountability and redress in the event of harm.

Often the focus is on the 'mean' of the outcome distribution to justify use and improve the validity of AI applications, however risk thinkers must assess the 'tails' of those same distributions, with their peripheral vision and creative minds, to be able to mitigate any unforeseen, disastrous consequences. It was with this mindset of imagination and exploration in which the Financial Industry Forum on Artificial Intelligence (FIFAI) had operated — a gathering of Canada's finest financial services experts in the application of AI, mixed with representatives from government bodies, and academia. We listened to leading AI experts and influential regulators from several countries, and then debated those learnings with the objective of safely harnessing AI potential to contribute significantly to Canada's economy.

This report summarizes those discussions and brings forth a framework that examines the challenges and opportunities for creating effective regulation. While in many instances participants voiced a range of views, one message was unequivocal and unanimous — necessity and desire for multidisciplinary, genuine, and effective collaboration. Bringing together interdisciplinary teams not only facilitates faster and seamless adoption of AI technology at an individual firm level, but also creates an opportunity for the broader ecosystem to jointly tackle common risks and challenges that have emerged in this space.

We are very grateful to all participants for sharing their knowledge, time, candor, and valuable contributions. We hope that this report contributes to the responsible and safe adoption of AI in Canadian financial services and influences a broader discussion on AI.

Angie Radiskovic
Assistant Superintendent and
Chief Strategy and Risk Officer
Office of the Superintendent of
Financial Institutions

Sonia Baxendale
President and Chief Executive Officer
Global Risk Institute

Executive Summary

Data availability and accessibility have improved dramatically, modeling techniques have taken a large step forward, and models are being applied to an increasing number of businesses across regulated financial institutions in Canada. Capabilities and usage have evolved faster than regulation.

OSFI partnered with the Global Risk Institute (GRI) to create a community of AI thought-leaders from academia, regulators, banks, insurers, pension plans, fintechs, and research centres. This group, called the Financial Industry Forum on Artificial Intelligence (FIFAI), advanced the conversation around appropriate safeguards and risk management in the use of Artificial Intelligence (AI) at Financial Institutions.

Ideas discussed throughout the workshop series to support safe AI development are grouped into Explainability, Data, Governance, and Ethics — the “EDGE” principles.

Explainability should be considered at the onset of model design and is driven by the use case and associated governance framework. Examples were provided where an explainable model would be selected over a higher performing opaque model, recognizing that modeling goals may be broader than strictly performance. In high impact use cases, there was discussion whether inherently explainable models should be used rather than relying on post-hoc explanations.

Financial institutions have long been working with **data**, but the integration of AI and the ensuing data sources and automated self-tuning algorithms into their operations have presented new challenges for managing and utilizing data. With new data sources and types, increased volume of data, and acceleration at which data is generated and utilized, it can be more challenging for financial institutions to integrate and standardize controls to manage data risk. This is especially true when data exists in silos within an organization or when it comes from different external sources. Improving the data used to train an algorithm will have a direct impact on model performance. Therefore, it is important for financial institutions to align their business and data strategies to ensure that they are collecting, managing, and analyzing the right data to support their goals. Good data governance can help ensure that data is accurate, consistent, and complete, which is crucial for the effective functioning of AI systems.

Governance has become an increasingly important topic and has matured to have the following properties: it should be holistic and encompass all levels of the organization, roles and responsibilities should be clear, it should include a well-defined risk appetite, and it should be flexible as an institution's adoption of AI matures. In addition, AI model governance specifically requires a multi-disciplinary approach to be effective. When governance becomes a rote exercise, focus drifts from understanding risks towards completing every element in the prescribed framework, regardless of risk or relevance.

The concept of **ethics** is very nuanced and naturally it is subjective. Ethical standards change over time and their codification into laws and regulations show the challenges and complexity of addressing AI Ethics. There is not a universal definition of fairness: what is perceived as fair depends strongly on the context as well as one's perspective. Within the realm of algorithmic fairness, there are different mathematical definitions, many are conflicting. Despite the legal perspective, complying with the law does not always mean that actions and outcomes are fair or perceived to be fair. There are many use cases where bias is the desired outcome, such as pricing policies or risk stratification. Data used for AI training can be the source of bias

and unfair outcomes. The current approach to addressing potential discriminatory bias is dubbed 'fairness through unawareness' where financial institutions do not use, and may not even collect, certain personal attributes in decision-making. This renders their models 'attribute blind' but may not be outcome-neutral (and it becomes difficult to test without the attributes). The societal expectation that financial institutions maintain high ethical standards continues to increase and there is real reputational risk and consequences when harm, actual or perceived, is done to customers. Organizations should maintain transparency, both internally and externally, through disclosure on how they ensure high ethical standards for their AI models.

Globally, regulators are striking a balance between regulation and innovation, that is, setting robust regulations while ensuring financial institutions continue to transform and remain competitive. The approach to regulating AI globally varies across jurisdictions, with institutions like the Bank of England adopting a principles-based approach while other jurisdictions, like the Monetary Authority of Singapore, provide more granular prescriptive guidance.

As OSFI looks to the future, an enhanced E-23 Enterprise Model Risk Management Guideline will be released for draft public consultation later in 2023. The insights and discussion from FIFAI are a testament to the need for collaboration and a multidisciplinary approach and have shown an appetite for continued dialogue in the Canadian financial services industry.

Introduction

There has been rapid growth in digitalization and usage of AI across the financial services industry. As the use of these technologies continues to evolve, current AI risk management frameworks must adapt to remain relevant, forward-looking, and responsive to industry needs. For the purposes of this report AI tools, models, applications, and systems will be used interchangeably.

In September 2020, the Office of the Superintendent of Financial Institutions (OSFI) published a discussion paper which identified core principles to manage risks associated with the use of AI by financial institutions. The goal of the paper was to comment on the implications of soundness, explainability and accountability with AI models. Since publication of this technology risk discussion paper, OSFI has led industry surveys and deep-dive studies with selected financial institutions along with bilateral exchanges with research centres and pertinent industry forums to better understand the uses and blind spots of AI.

OSFI recognized that an industry approach to the safe adoption of AI in financial services was needed and partnered with the Global Risk Institute to host the Financial Industry Forum on Artificial Intelligence, featuring four workshops. The workshop series included events with participants from both public and private sectors present. Selected participants were invited to share their perspectives on the evolution of AI, emerging themes, use cases, and challenges with regards to adopting AI, its development, deployment, and use in financial services. Participants brought their experiences from other sectors to discuss what can be leveraged within the financial services industry, along with best practices on addressing the challenges, mitigating, and managing the associated model risks, and model governance to ensure responsible AI adoption.

In the first workshop, participants identified the areas of greatest importance to AI models. These topics: explainability, data, governance, and ethics would form the basis of discussion for subsequent workshops and the structure of this report. The four themes are collectively referred to as the "EDGE" principles.

The EDGE Principles

Explainability enables financial institutions to deepen trust with their customers. When customers understand the reason for decisions, they become empowered to work towards their financial goals.

Governance supports the realization of AI's potential by ensuring that the financial institution has the right culture, tools, and frameworks available to support the AI lifecycle.

Data leveraged by AI allows financial institutions to provide targeted and tailored products and services to their customers, improve fraud detection, enhance risk analysis and management, boost operational efficiency, and improve decision making.

Ethics encourages financial institutions to consider broader societal impacts of their AI systems and make a conscious choice of what role they would like to play in shaping the world around them.

This report delves into the key takeaways from a forum exploring the integration of AI in the financial services sector. Drawing from the discussions among participants, it sheds light on areas of agreement and disagreement. Furthermore, it encompasses the perspectives of keynote speakers, offering a wider view on each theme of FIFAI. With the goal of helping practitioners effectively manage the challenges and potential of AI in the sector, as well as the best approaches to adoption and management of the technology, this report aspires to promote stronger risk management practices.

Explainability

One of the most prominent and persistent challenges in utilizing AI techniques is the potential difficulty in explaining how those models reach a conclusion.

Explainability permits the examination of the theory, the data, the methodology, and other key foundational aspects of an approach, prior to verification of its performance, to confirm that the model is fit for purpose.

Explainability of AI models can facilitate assessment of fairness and generally aid in protecting against bias and discrimination. We are now beginning to see regulators formally address this issue of explainability. In addition, there has been an increase in research focused on explainability in AI.

The forum addressed the following key questions on explainability:

- What levels of explainability might AI systems have?
- What factors should determine an appropriate level of explainability for a particular application?
- What are the approaches to achieve explainability and what are the associated risks?
- How does explainability connect to a more general concept of transparency?
- What is the role of explainability in building trust?

> **"One of the key things that explainability enables is ensuring that the right decision is being made for the right reason."**
>
> Alexander Wong, Professor and Canada Research Chair, University of Waterloo

Explainability Dimensions

The degree of explanation required for a model should be considered at the onset of model selection and design and is driven by the use case and associated governance framework. For example, explainability could be helpful to data scientists for easier debugging and better identification of ways to improve performance and robustness of AI models. Explainability can help business owners understand and better manage risks that stem from AI tools, and also help regulators certify compliance. Customers may require explanations to understand why a certain decision was made, and how the customer can change their behaviour to influence future decisions. At a broader model level, explainability helps facilitate financial institutions to assess whether the AI-driven decisions align with their corporate values and contribute to responsible use of AI.

Appropriate Level of Explainability

Levels of explainability reflect the degree to which we understand how a model arrives at its conclusions. Models that are completely transparent have a high level of explainability compared to less transparent techniques that have a low level of explainability. An explanation of outcomes can still be provided for models with a low level of explainability via post-hoc analysis techniques, at both the local and global scope.

There is an ongoing conversation on the level of model explainability that is appropriate for a given use case and stakeholder group. One perspective on explainability is its importance in most use cases for ensuring that AI models are used ethically. For instance, the European Union's (EU) General Data Protection Regulation (GDPR) requires a "right to explanation" for users on all decisions made by automated systems. While the EU is one of the earliest jurisdictions to issue such regulation, there is an interest globally to develop similar guidance. Many forum participants advocated that they would be comfortable using less explainable black-box models to make decisions so long as there were sufficient controls to prevent the model from performing outside of a pre-defined boundary.

All forum participants agreed that the appropriate level of explainability required will depend on several factors, including:

- **What needs to be explained?** Some approaches provide the ability to explain the significance of each variable for a particular prediction (e.g., feature importance as in Shapley values). On the other hand, one can strive for higher levels of explainability to understand

how the entire model works (e.g., using interpretable models such as decision trees). The role of AI in the final decision could be another aspect included in the explanation as it could be important to know if the decision was made directly by an AI system or if it was an AI-assisted decision.

- **Who needs the explanation?** Levels of explanation could differ depending on the recipient (data scientists, business owners, regulators, or customers). Indeed, an explanation that would suffice for a customer may be insufficient for a data scientist because the two parties need explainability for different reasons.
- **Is this a high-materiality use case?** The need for explanations is less critical for chatbots or AI models used for marketing than for AI models used to make credit decisions or measure capital. It is commonly accepted that higher levels of explainability are required for high-materiality applications.
- **How complex is the model?** Some models could be complex to the extent that little can be explained about the model and outcomes which may imply that those models be considered inappropriate for certain use cases.

Explainable AI

While some models can be explainable by design, some are deemed black-box models and need additional techniques to help understand the model outcomes. Those additional techniques are referred to as post-hoc techniques. Models that are explainable by design are also termed inherently interpretable or explainable models, their inner mechanisms can be inherently analyzed and understood.

The use of techniques to help understand or interpret model outcomes forms the area termed *Explainable AI*. Such technique is a separate model that aims to replicate the behaviour of the black-box AI model. In other words, when we say “Explainable AI” it is implicit that there are two models: the black-box model providing decisions that should be explained and the inherently explainable model that is designed to replicate the behaviour of the black-box model.

Explanations without Biases

Researchers from the University of California (Irvine) and Harvard University demonstrated that LIME and SHAP explanation techniques could be unreliable. Using extensive evaluation with multiple real-world datasets, they demonstrated how biased classifiers crafted by their framework can easily fool LIME and SHAP into generating innocuous explanations which do not reflect the underlying biases.[1]

Approaches to Explainability

The discussions on Explainable AI also covered different types of explanations, distinguishing between "understanding a particular decision" (local explanation) versus "understanding a model" (global explanation). In case of local explanations, a data scientist might be able to provide an explanation for why a particular decision (e.g., loan adjudication) was made, but it might not imply that the data scientist knows the broader inner workings of the model for all decisions.

Some techniques are used to provide local explanations while some are used to provide global explanations. For instance, to explain the contribution of each input feature to a particular decision made by a black-box model one can employ Local Interpretable Model-agnostic Explanations (LIME) or Shapley additive explanations (SHAP).

While these techniques provide insights into model outcomes, they are not entirely flawless. Although there is a common opinion that explainability can help in finding biases, one should know that LIME and SHAP could leave biases undetected (i.e., biases are present, but not reflected in explanations).

In addition, it is worthwhile to note that some techniques or models utilized to explain a "black-box" model can result in different explanations for the same prediction. To address the issues of post-hoc techniques or explanation models, it could be a good practice to validate the models used to explain "black-box" models as well as the corresponding "black-box" models.

Interpretability-Performance Trade-off

The possibility of a trade-off between interpretability and performance was explored during the forum. This particularly applies to some complex and powerful models, such as neural networks, which could yield better performance but lack high levels of explainability present in inherently interpretable models.

Forum participants were divided in their response to a trade-off between

interpretability and performance. While dependent on use case, some felt that the best approach was to use inherently interpretable models and focus time and investment on improvements in data. Several forum participants agreed that model performance from an inherently interpretable model could be comparable to model performance from black-box techniques given significant enhancements (quality, availability) to data. Other participants were comfortable relying on post-hoc explanations provided for their high performing and low inherently interpretable models.

Explainability and Performance

There is a widespread belief that more complex models are more accurate, meaning that a complicated black box is necessary for top predictive performance. However, Cynthia Rudin, Professor at Duke University, noted that this is often not true... Even for applications such as computer vision, where deep learning has major performance gains, and where interpretability is much more difficult to define, some forms of interpretability can be imbued directly into the models without losing accuracy.[2]

Disclosure

Disclosure is important for financial institutions because it helps promote transparency and accountability. Financial institutions are required to disclose adequate and relevant information to investors, regulators, and the public to help them understand the institution's financial health and performance. This allows investors to make informed decisions about whether to invest in the institution and helps regulators to ensure that the institution is complying with laws and regulations. Additionally, disclosure helps to build trust between financial institutions and their customers, as it demonstrates that the institution is open and honest about its operations.

Disclosure of adequate and relevant information on AI models is also important to financial institution customers. When a decision is made about a customer, best practice (and where required by law) would necessitate an explanation be provided on how such decision was made. The challenge is for model developers to design these systems in a way that can satisfy accuracy and disclosure goals. Similar to the earlier discussion on the appropriate levels of explainability, participants agreed that the details included in disclosures depend on the use case and its risk and materiality.

It is also generally agreed that customers should be informed when engaging with AI. Canada's Treasury Board Secretariat through its Directive on Automated Decision-Making has established requirements that customers be provided notice that the decision rendered will be undertaken in whole or in part by an Automated Decision System.

Properties of Good Disclosure

Regarding the properties that a good disclosure should have, it was suggested that disclosure be concise, simple, relevant, intuitive, and practical. In addition, disclosure should be written in plain language that is understandable to users without any AI expertise, and accompanied by examples and recommendations, where appropriate. Finally, they should also be presented in a logical and organized manner and be tailored to the specific needs and background of the audience.

Risks of Disclosure

As highlighted, there are many advantages for financial institutions to provide AI related disclosures. However, care should be taken when deciding what information is disclosed. The following factors should be considered to mitigate unintended consequences of excessive disclosure:

- **Cyber security:** Disclosing seemingly innocuous information could undermine the cyber security of financial institutions.
- **AI integrity process:** Disclosure should not compromise AI process. For instance, poisoning attacks in adversarial AI aim to influence the data used in training or re-training the algorithm. Contaminated data is fed into the algorithm and causes the machine to learn the wrong way. This is especially important if the data used by the algorithms is publicly available and susceptible to compromise.
- **Competitive advantage:** When an organization shares too much information about its AI products, services, strategies, or other proprietary information, it can potentially lose its competitive advantage.

In general, companies should strike a balance between disclosing enough information to satisfy customers and investors while keeping sensitive information confidential to maintain their competitive edge.

Third-Party Disclosure

Financial institutions are not able to adequately disclose to their customers if they do not have full visibility over their third-party models since those providers treat their products as proprietary in order to protect their intellectual property (IP).

Forum participants discussed potential solutions to address explainability and disclosures related to third-party AI models. One solution was to incorporate explainability requirements as part of third-party tools'

procurement process. Some forum participants thought such solution would be difficult to implement as third-party providers would disagree with those terms, unless that was industry expectation, set by all financial institutions. Another solution for third-party AI model explainability was the certification of third-party models by an independent body. The success of the proposed solution cannot be ascertained as there isn't an established framework for third-party model certification. In addition, given the bespoke nature of AI models and for different use cases, there are challenges with providing third parties blanket endorsements for their products. There is also a risk that such an independent body could create a monopoly stifling innovation or dilute the quality of third-party certifications. A third proposed solution was to disclose information related to the third-party application directly to the regulator on a case-by-case basis to allow the regulator to verify the integrity of such application.

Instead of a compulsory certification, forum participants agreed that consistent industry-wide standards could be adopted for models supplied by third parties (like an ISO accreditation). Voluntary adoption of certifications and standards would not guarantee regulatory approval.

A few participants at the forum advised against using third-party AI models whenever possible in favour of internal development to avoid issues related to explainability.

Explainability and Trust

The relationship between explainability and trust (i.e., a firm belief in reliability of AI model) is not as obvious as it might seem. Oftentimes, explainability is broadly perceived as a tool to build trust. Sole knowledge of how the model makes decisions does not necessarily lead to complete trust in the model as there are other aspects that may be considered, such as model accuracy and absence of bias.

However, there are some instances where explainability can inhibit trust, such as when the customer does not agree with the explanation for a given decision. An explanation that is difficult to understand could also decrease overall trust in the AI system.

Explainability, together with disclosure at the right levels and to the right audience, is one of many factors that contribute to developing trust between a financial institution and its customers. Inevitably, increasing trust in AI enables further use and innovation.

Data

Data is a crucial resource for the development and implementation of AI technology.

As more data becomes available, AI applications are able to improve and expand. Financial institutions have been able to leverage data to reap benefits such as tailoring customer service, improving fraud detection, and boosting operational efficiency, however, the integration of AI into their operations has further highlighted challenges for managing and utilizing data.

One of the main issues is ensuring data quality, which encompasses data being accurate, reliable, complete, representative, consistent, and compliant with relevant regulations. Another one is privacy, as financial institutions must take steps to protect sensitive personal and financial information. Additionally, aligning data strategy with business strategy has become more complex. There is increased importance in achieving sound data governance in order to address those issues.

> **“Everyone talks about models, models, models. We need to focus on proper data analysis first… shift [the] mindset to [data] stewardship.”**
>
> Ima Okonny, Chief Data Officer at Employment and Social Development Canada (ESDC)

The forum addressed the following key questions on data:

> What is the impact of a varied dataset on data quality?
>
> What challenges has AI introduced for data governance?
>
> How can AI-specific risks arising from third-party exposure be addressed?
>
> What are the complexities in aligning data and business strategies?

AI can be trained on a variety of types of data, including:

Structured data: data that has specific structure. Examples of structured data include financial transactions, customer information, and inventory data

Unstructured data: data that is not organized in a predefined data model or structure, such as free-form text, images, and video. Examples of unstructured data include social media posts, emails, customer reviews, board reports.

Synthetic data: data that is artificially generated. This type of data can be used to supplement or replace empirical data for training AI models.

Data Characteristics and Associated Challenges

Data used for AI training and development exhibit a number of characteristics which when leveraged by AI present a wide range of possibilities. However, these characteristics can make it more challenging for financial institutions to integrate and standardize data as well as manage data risk. They are discussed below:

- **Data Volume and Variety:** AI models can process large amounts of data and can handle structured and unstructured data, thus, increasing the challenges to ensure continuous data quality and integrity.
- **Data Versioning:** AI models can be iterative and trained on multiple versions of data. This feature can make it particularly challenging to keep track of the version of data that was used to train a particular model and how it may have evolved over time.
- **Data Agility**: In addition to the speed at which data is generated and leveraged by AI models, some real-time systems react to live data feeds. Strong governance is needed in these use cases with a human-in-the-loop and offline models that can be activated should the real-time system respond in unexpected ways to live data inputs.

Data Quality

Data quality is critical for AI applications. The task of maintaining high data quality has increased in difficulty as outlined by the points above. In addition, there are other issues that further exacerbate the difficulty in ascertaining the quality of data, they include:

- **Inconsistency:** Data can be highly inconsistent, with varying formats, structure, and levels of detail. This can make it difficult to identify patterns or trends in the data.
- **Noise:** Typos, grammatical errors, and irrelevant information often found in data can make it difficult to extract useful insights.
- **Lack of context:** Without proper context, understanding the meaning and significance of some data can become challenging. For instance, for a sentence in a customer feedback form “I am not happy with the service” , it is not clear what service is implied.
- **Quality of sources:** Data can come from a variety of sources, such as social media, customer feedback, and news articles, which can have varying quality and reliability.

- **Dual meaning:** Some types of data can have vague interpretation or be understood differently depending on their business use. For example, notional amount in financial transactions or derivatives. [3]

To overcome these issues, data scientists and engineers may undertake data exploration, data cleaning, data validation, and data integration. However, it is worth noting that even with these approaches, it can still be difficult to assure sound quality of data.

Forum participants also noted that although synthetic data can be used to address certain issues of data quality (e.g., bias, fairness, imbalance), the research in this area is still in its early stages. Moreover, it is important to emphasize that synthetic data can have its own limitations such as not being able to fully replicate the real-world complexity and variability, which could lead to underperforming AI models.

Synthetic Data

As highlighted by Stuart Davis, EVP Financial Crimes Risk Management at Scotiabank, synthetic data is a big area of emergence. Due to privacy and other restrictions, financial institutions cannot always give real data to partnering institutions and synthetic data could be a good alternative.

Data Aggregation

When collecting and combining data of different formats and from multiple sources, it is important to recognize the potential inconsistencies and errors that may arise. For instance, a crucial step in many supervised learning applications is labeling data. With different people or data providers labeling data, there is a chance that they may use different standards or definitions, resulting in inconsistencies in the data that pose a challenge when such datasets are aggregated. This can lead to poor performance of the AI system, as it may not be able to properly learn from the inconsistent or incorrect labels. To avoid these issues, it is recommended to have clear guidelines and protocols in for data labeling, lineage, and management.

Data Governance

Good data governance can help ensure that data is accurate, consistent, safe, and complete, which is crucial for the effective functioning of AI systems. Data governance is critical for financial institutions considering the sensitive and confidential nature of financial and customer data. Forum participants explored some aspects associated with data governance.

Data Ownership

Data ownership issues could present certain challenges for organizations as datasets necessary for building an AI model may be sourced from different business areas. Seeking permission from the different business owners to use their data, although necessary, could slow down the AI development process. While each business area is accountable for their own data, it is the team that builds the model that should ensure that the data used to build the model is correct and complete.

Data in Silos

During his presentation at the forum, Andrew Moore, Director of Google Cloud AI, discussed Google's work on an Anti-Money Laundering tool which turned out be a challenging task to accomplish because of data availability. He noted that in the financial services industry, data is usually living in silos across an organization and, while building the tool, Google had to integrate thousands of databases. Thus, deploying AI tools that worked well in tech applications could be far more challenging in finance due to the data issues.

Data Privacy and Security

Large amounts of data are collected by organizations for their AI models to leverage and generate insights, however, some of the data may be personal or sensitive. Where personal or sensitive data is present, data privacy and security are essential, though risks of data not protected against leakage or unauthorized access continue to increase. Sound data governance mitigates these risks.

Regional Data Limitations

Of particular importance is the scalability of AI-based solutions across different geographies. One of the challenges is the use of protected variables because there is no consensus across jurisdictions on the suite of variables deemed as protected.[4] Issues around protected variables could also create a burden for data governance (e.g., tracking, managing, granting permissions) and slow down model development process.

Another relevant challenge for financial institutions that span business across different regions could be the interaction between different systems or processes across jurisdictions, including legacy infrastructure and even timing of data.

Data-Centric Approach

One way to improve the performance of both AI-applications and traditional models is to continuously improve the data used to train those models, also known as a data-centric approach. Rather than solely focusing on algorithm iteration and retraining to improve performance, incorporating a data-centric approach maximizes the performance potential of the model. Sound data governance is necessary to adopt a data-centric approach to AI model development.

Data Literacy

Building a strong data literacy culture has been also identified as vital for organizations that actively deploy and use AI. Organization-wide awareness of the various risks that stem from inadequate use of data is essential with widespread adoption of AI, thus, organizations should consider ongoing training activities for their employees on a broad range of aspects related to data.

Third-Party Data

The value from AI is dependent on the quality and quantity of data utilized. Organizations need data beyond those generated from their business activities, therefore, have to rely on third parties. Forum participants explored aspects related to the use of third-party data.

Data Collection

With situations where data is collected by a third party and used by a financial institution, there could be a lack of clarity on the extent to which the financial institution is responsible if the third party experiences a data breach. In such cases, even if the third party is accountable, the financial institution could still be exposed to reputational risk.

Considering the variety of data types that can be used by AI models, special attention should be given to the aggregation process for data that is sourced from different third parties in order to prevent inconsistencies and errors within the aggregated data.

It could also be difficult for organizations to verify that data was collected properly by the third party, thus, requirements to assure appropriate data collection by a third party should be put in place. For example, it could be required to confirm that the data vendor obtained consent from the data subjects before supplying the data to other third parties or the financial institution.

Forum participants generally agreed that regulatory rules should be very clear about what types of third-party data can be used by financial institutions and the level of due diligence required in each case. When data is obtained from third parties, certain guardrails are essential to ensure governance for data provenance, data lineage, and data quality. A suggestion was made to consider adopting a “nutritional label” approach to data, where key information is disclosed on the dataset such as when it was collected, how it was collected, what was its intended use upon collection, etc.[5]

Data Sharing

AI cannot be considered in isolation from increasing digitization of our society. The Internet-of-Things (IoT) and Open Banking (OB) are the two prominent developments that were extensively discussed during the forum. It was noted that on the one hand, IoT and OB could exacerbate the issues of data privacy because more market actors can potentially access to data. In this respect, accountability could become more challenging with OB when data from one

party could be used by another party. On the other hand, data sharing through OB could potentially help in detecting bias and improving overall performance of AI models due to larger sets of data available to train AI models.

Alignment of Data and Business Strategies

Aligning data strategies with business strategies is essential for organizations to effectively leverage their data assets to drive business outcomes. A data strategy outlines an organization's approach to managing and leveraging data, while a business strategy outlines the organization's overall goals and objectives. When data strategies are aligned with business strategies, organizations can ensure that they are collecting, managing, and analyzing the right data to support their business goals.

Additionally, as AI brings new capabilities and opportunities to an organization, such as automating processes, identifying new insights, and creating new products and services, the overall business strategy may have to be adapted to the capabilities of AI.

While the benefits were recognized, it was also acknowledged that it could be challenging for financial institutions to ensure that their data strategy accounts for the specific requirements of AI and that it supports the organization's overall business strategy. It was discussed at the forum that it is not always evident to quantify the benefit of investing in data strategy, especially when compared with other immediately profitable business projects, leading to possible sub-prioritization of strategic data-related projects.

Governance

By taking a proactive approach to governance, financial institutions demonstrate their commitment to the responsible use of AI and build trust with their customers and stakeholders.

A robust governance framework promotes a culture of responsibility and accountability around the use of AI within an organization. This allows financial institutions to fully realize the benefits made possible by AI while avoiding harm to customers and the broader society.

The forum addressed the following key questions on governance:

- What constitutes good governance for AI models?
- Are the existing Model Risk Management approaches sufficient?
- What challenges are institutions facing on implementing governance frameworks for AI?
- What tools or best practices can be used to mitigate the risks and challenges of governance

"AI use at financial institutions is changing fast. AI is a business enabler and the benefits can be seen in various ways, from accelerating internal productivity to driving growth, and can be linked to corporate performance. It brings up the question of: what is the right balance between risk and innovation from both a business and a regulatory perspective?"

Donna Bales, Principal Research Director in the CIO Practice at Info-Tech Research Group, Founder of Canadian Regulatory Technology Association (CRTA)

An AI Governance Framework

While governance includes oversight, it is a broader concept. Governance refers to the structures, systems, and practices an organization has in place for decision-making, accountability, control, risk monitoring and mitigation, and performance reporting.[6]

Although financial institutions have implemented governance practices with varying degrees of sophistication, the increasing use of AI techniques has sparked discussion on what governance framework changes are needed to support adequate control.

Model risk management came into focus following the global financial crisis of 2008 with the introduction of regulatory guidelines from the US Fed with SR11-7 (2011),[7] OSFI with E-23 (2017),[8] among others. These guidelines highlight the importance of model risk management as a component of good governance and elevate its focus within a broader enterprise risk management framework.

AI models pose many of the same risks as traditional models and often exist within an ecosystem that interacts with other established risk and governance functions.

Forum participants highlighted characteristics desirable for good governance of AI at financial institutions:

- **It should be holistic and encompass all levels of the organization.** It is important for all internal stakeholders to understand AI fundamentals. This can help in managing risks that stem from AI-based applications. Senior management up to board level should understand the benefits, risks, and limitations of AI model use to ensure appropriate decision-making. Enterprise-wide processes and frameworks are instrumental for good governance.
- **Roles and responsibilities should be clear and well-articulated.** Accountability with respect to data, AI models, and outcomes as well as the structure of approvals with respect to different risks need to be defined. Clear articulation of mandates for those different groups prevents gaps in risk management.
- **It should include a well-defined risk appetite.** Financial institutions should define or update their risk appetite taking into consideration the increased risks that arise from the use of AI including cyber or third-party risk.

- **It should reflect the risk of use cases.** This includes a cultural transition within financial institutions from a rule-based to a risk-based approach, permitting risks to be taken in accordance with the enterprise risk appetite statement. The level of due diligence applied to the AI use case is commensurate with the risk.
- **It should be flexible as a financial institution's adoption of AI matures.** Financial institutions that successfully adapt existing governance frameworks for AI models will embrace flexibility and agility, pivoting where needed as new techniques and risks emerge.

The Treasury Board Secretariat (TBS) Algorithmic Impact Assessment Tool

An example of a tool for good governance, was presented by Benoit Deshaies, Director, Data and Artificial Intelligence, TBS. The TBS AIA tool is a mandatory assessment tool that applies to all Government of Canada departments to assess automated decisions on a range of topics. It was developed using a collaborative approach, has a well-defined scope and application, is risk-based, and it implements the principle of proportionality, as it determines a score based on the impact to customer. Risk mitigation practices are then translated into specific requirements of governance depending on the level of impact. It is a self-assessment tool that is supported by a peer review process for automated systems that may require it.[9]

Evolving Existing Governance Frameworks for AI

Through its AI Public Private Forum, the Bank of England found that "existing governance frameworks and structures provide a good starting point for AI models and systems" . The Bank of England concluded that model governance should align to the risk and materiality of the use-case, with special consideration given to governance vulnerabilities exacerbated by AI models.

Participants at FIFAI broadly agreed with the Bank of England's conclusion that extending existing governance frameworks was a better approach than developing a suite of new AI-specific processes and procedures. Financial institutions are at different levels of maturity in their adoption of AI and even in their implementation of governance frameworks. Developing a robust governance framework inclusive of AI might involve a significant culture change as more areas of financial institutions leverage AI techniques.

According to the Institute of International Finance (IIF) Machine Learning Governance Summary Report (2020), there are differences in approaches across geographies. Financial institutions in Canada are placing emphasis on enhancing existing frameworks. [10]

FIFAI participants reflected on the key challenges with applying and adapting existing governance frameworks. As with any governance framework, a key risk is compliance becoming a "check the box" exercise, as explained in "Model Risk Management Lessons Learned: Tracing Issues from the Pandemic to the Great Recession" . [11] When governance becomes a rote exercise, focus drifts from understanding risks towards completing every element in the prescribed framework, regardless of risk or relevance.

Model Risk Rating

A governance framework normally outlines the risk grading of models to account for the nuances in risk arising from model use. Institutions have developed different approaches to rate or grade model risk. Generally, these approaches consider materiality, financial impact, complexity of the methodology, complexity of infrastructure, etc. It was discussed at the forum that an important aspect to consider for AI models should be customer impact as it could lead to reputational risk. This would mean considering not only the financial significance of an AI model but also the potential effect on customers. This approach takes a

comprehensive view of the risks associated with AI models by looking at both financial and non-financial aspects. Furthermore, this will be of relevance with the adoption and compliance with the Artificial Intelligence and Data Act of the forthcoming Canadian Bill C-27.

Model Inventory

Traditionally, institutions with mature governance frameworks maintain a model inventory which could be expanded to incorporate all AI models. A sophisticated approach identifies the interrelationships between AI models to reflect when outputs from one model are inputs to another or when a model is used to explain the outcomes of another. The risk rating and the model inventory are tools that can help keep track of the governance efforts required per model on a risk-based approach.

Approaches to AI Governance

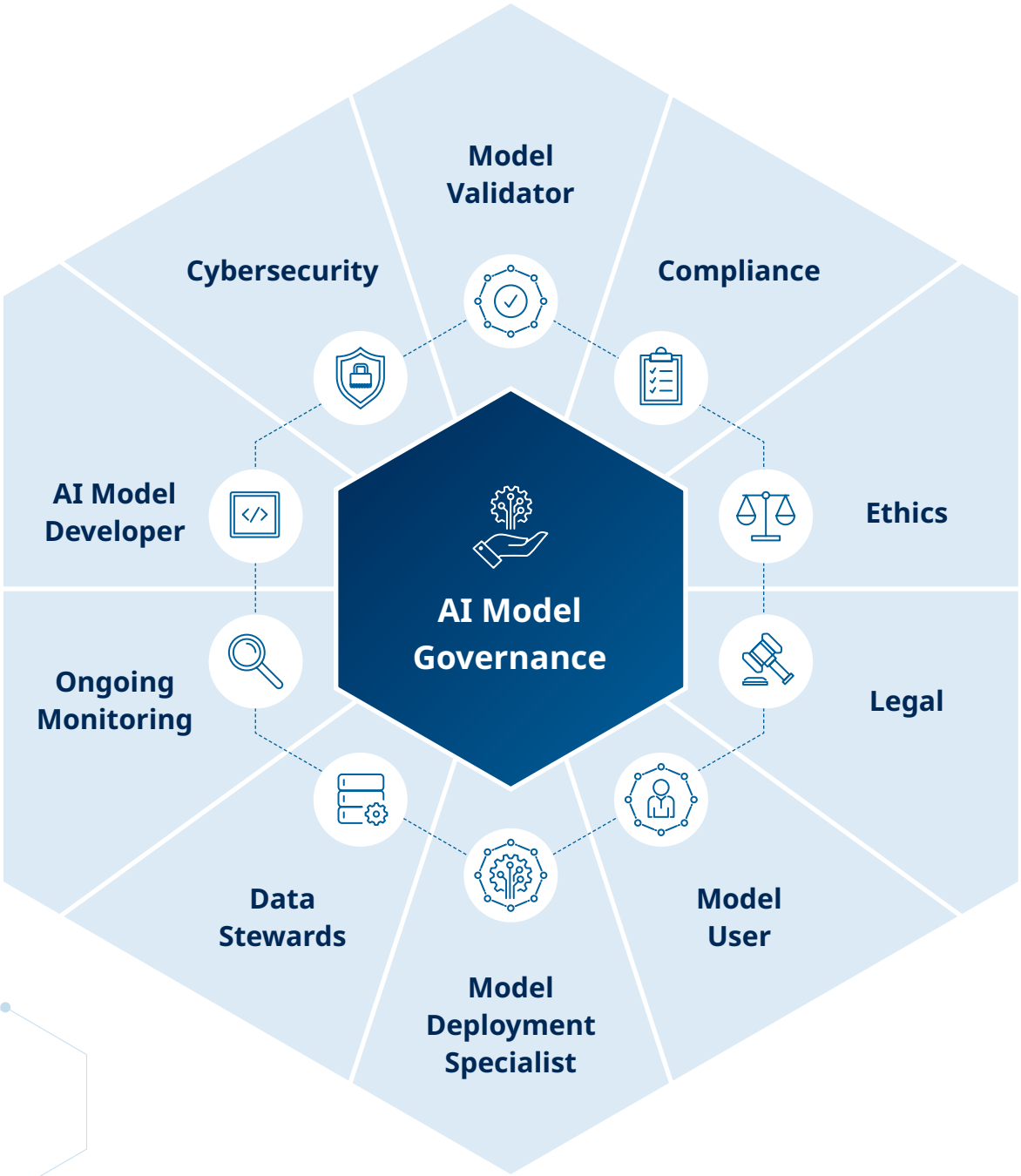
The exact composition of an AI governance structure will vary between financial institutions depending on factors such as the institution size and the sophistication of its enterprise governance framework. The IIF survey referenced above shows that 35% of global financial institutions have established a centralized specialist team, such as an AI Governance Council, that assists with alignment of organizational goals, assessing AI use case value and prioritization, and approving the AI models prior to deployment. For smaller institutions with low-risk use cases, accountability may be retained within the business unit that uses an AI tool. A third, hybrid approach exists between a centralized and federated governance structure, where a coordinating team brings together the different stakeholders relevant to develop and govern the AI use case.

Multidisciplinary Approach

AI model governance can be more complex than traditional model governance as it requires a multi-disciplinary approach to be effective. Traditionally, model governance is subject to a model review committee comprised of senior leadership representing the model development, validation, and user teams. Given the increased scope, scale and complexity of some AI use cases, the sufficiency of this stakeholder group was discussed. For example, some forum participants felt that accountability for ethical issues be explicitly defined in the governance process. The use of AI models may also introduce legal considerations which should form part of the governance process. Compliance and legal teams could be key stakeholders in the development of AI models.

Figure 1 presents an example of stakeholder groups that could be integrated as part of a model governance process.

Figure 1
Multidisciplinary Team for AI Governance



Accountability

In considering accountability, forum participants discussed the need to integrate data and model governance. Participants acknowledged the distinctness of data and model accountability as the skillsets, tools and processes needed for model and data governance differ. However, participants agreed that effective data governance is foundational and feeds directly into model governance, hence there should be strong linkage between them. Model owners and model developers need to ensure the data used for the model is adequate and has appropriate controls to prevent risks like bias, unfair outputs, over or underfitting and lack of representativeness. This can only be achieved through proper data governance and communication between the applicable stakeholder teams.

Skills, Culture, and other Challenges

Forum participants discussed challenges with implementing a strong governance framework for AI models, they are presented below along with implications. Overall, effectively integrating AI tools into the financial services industry requires a combination of technical expertise, organizational change, and cultural shift.

Resource Competition

A competitive labor market has impacted the ability to attract and retain AI expertise within the financial industry. This is further compounded by the limited availability of individuals with a blend of AI and domain knowledge expertise.

Independence of Development, Validation and Audit Functionss

When the appropriate talent is attracted to financial institutions, they are often hired into a first-line (model development) function. This leads to an AI skill gap between development

and validation functions as well as between the development and audit functions. Although this is normal when new technology is nascent, it can challenge the effectiveness of the governance process.

Another aspect of this challenge is the use of an "agile" approach to model development. Participants at the forum were split regarding the inclusion of model validation team at the start of the model development process. Benefits cited were early identification of issues while weakened governance was highlighted as a challenge.

Reduced Human Oversight

The possible autonomous decision-making by AI systems could lead to reduced human oversight on key governance aspects. Changes in control framework such as, more frequent monitoring, are needed to mitigate some of the risks from automated decision-making.

Third-Party Solutions

AI systems can rely on vendor models, tools, or third-party data where lack of transparency can create governance challenges.

Open-Source

Risk that stems from open-source data and tools is different from third-party risk because there is no contractual agreement. Some participants felt that open-source code offered an advantage over code sold by a third-party provider. Proponents of open-source code highlighted its full transparency, and review by many users. Concerns with open-source code were related to accountability, where the onus is placed on the institution to review and assure soundness of the code being used rather than such expectations placed on a third party.

The Way Forward

Forum participants discussed the "way forward" for AI governance in the financial services industry and came up with the following key areas:

Tools and Technology Governance

Robust AI governance includes the ecosystem of tools and technology in which the model is developed, deployed, and monitored. This includes tools designed specifically to support the governance of AI models, for instance, Machine Learning Operations (MLOps[12]) which can help in automating aspects of the model life cycle such as development and monitoring. When implemented effectively, MLOps creates standardization and consistency by embedding existing data and model governance frameworks into its processes. MLOps should be viewed as a tool to support governance and not as a replacement. When AI tools are used directly to support governance, it remains important to maintain a human-in-the-loop to identify blind spots and gaps in governance.

Third-Party Governance

Discussions by forum participants led to an agreement that existing third-party risk management frameworks could be a good starting point to address AI-specific risks arising from the third-party exposure. There should be similar governance expectations between internally developed and vendor-provided solutions. To solve the "intellectual property" challenge of third parties, the forum participants explored a number of options.

With the industry trending towards open-source technologies, forum participants discussed the governance needed for open-source packages/libraries used to develop and implement AI models. Forum participants agreed that governance is needed for open-source code commensurate to the risk of the particular use case. In lower risk use cases and, where permitted by the enterprise risk appetite, financial institutions may use open-source code without a rigorous review process. In higher risk use cases, financial institutions should have an independent review of its open-source code.

Organizational Aspects

Financial institutions may need to create new roles or reorganize existing teams to effectively leverage the skills and expertise of data scientists and other technology professionals. Additionally, organizations may need to provide training and development opportunities to help both business and technology employees adapt to new technologies and workflows.

Skillset and Education

Some financial institutions are proactively addressing this skill gap through internal training. At a basic level, rotational programs are encouraged between lines of defense to facilitate the flow of talent and education across all areas of the organization. Some financial institutions are implementing AI-specific training programs that include law (for example, privacy, governance, human rights), cyber security, and domain specific knowledge. Institutions with an emphasis on education are better positioned to understand and manage the risks that stem from AI-based applications.

Collaboration

At an industry level, participants suggested continuous forms of collaboration, creating a broad community of practice where institutions could have opportunities to share best practices. Continuous dialogue and collaboration between different stakeholders like academia, industry and regulators could help advance innovation through knowledge sharing.

Ethics

Ethics in business are the moral principles and values that govern the way an organization makes decisions. Ethical principles include aspects such as right to recourse, fairness, and privacy, and are weaved into organizations' code of conduct and values. As the industry focuses more on concepts such as responsible investment, there is an increasing need to demonstrate a commitment to ethics in decision-making.

It was discussed at the forum that the concept of ethics encompasses a range of nuances. First, ethical principles and values can be relative, thus, impacting the way organizations in various jurisdictions address ethics. Moreover, the implementation of ethical standards may vary based on the specific application within an organization. For instance, while the use of travel history for fraud detection system could be considered appropriate due to its potential in identifying suspicious or fraudulent activity, its use to assess creditworthiness may be viewed as unethical as it may not be directly relevant to an individual's creditworthiness and may potentially discriminate against certain individuals based on their travel history. Finally, ethical standards could change over time. As an example, there is a much greater focus on Environmental, Social, and Governance (ESG) today compared to a few decades ago.

Intersection of AI and Ethics

David Leslie, Director of Ethics and Responsible Innovation Research at the Alan Turing Institute, defined AI Ethics as a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies.[13]

> **"We do not want technology that subverts our values such as the right to challenge a decision. Issues are created when we are led by technology rather than our values."**
>
> Carole Piovesan, Managing Partner at INQ Law

Various aspects of ethics within the context of AI in the financial services industry were discussed at the forum.

The forum addressed the following key questions on ethics:

> How does the relationship between ethics and law apply to AI?
>
> What are the different views on regulatory guidance for AI Ethics?
>
> What are the challenges in addressing AI ethics, and how can these challenges be overcome?
>
> What is the universal definition of fairness?
>
> Is a "biased model" necessarily a bad thing?

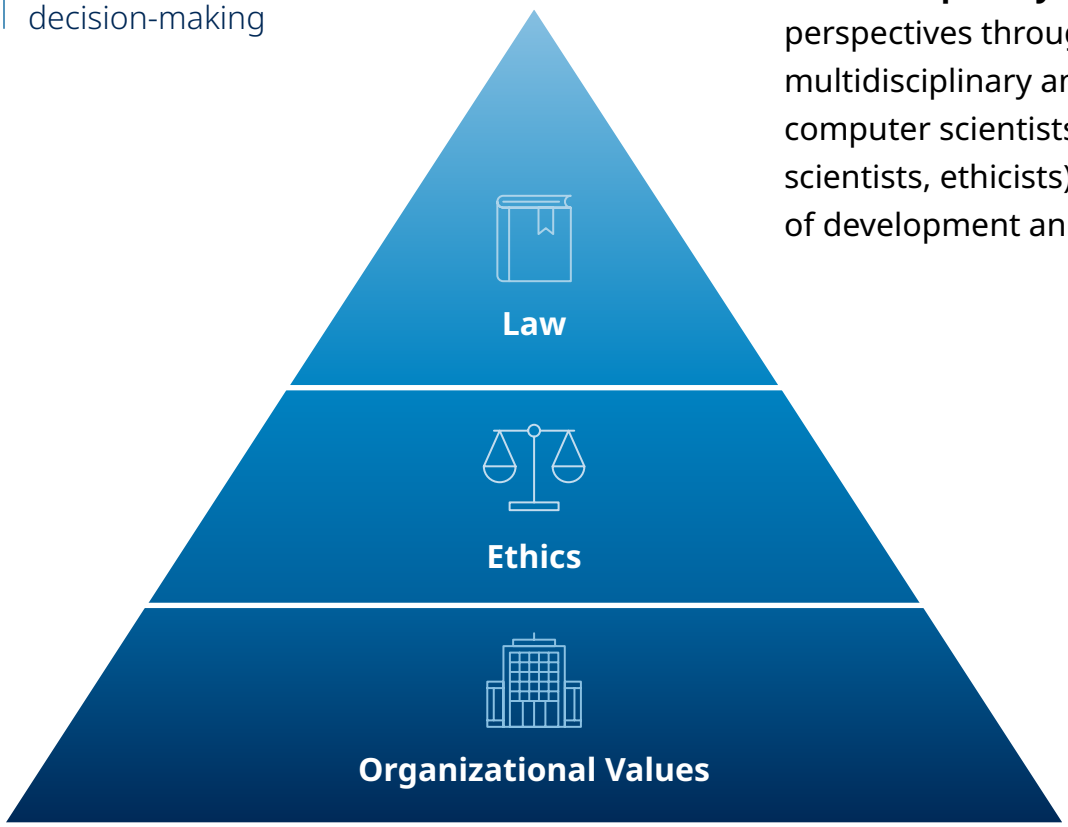
Legal, Policy and Regulatory Implications

The exploration of ethics cannot be made without considering the legal standards. Though related, law is different from ethics. Legal standards are set by the government, whereas ethical standards are based on principles and values that may go beyond what is legally required. As a result, it is possible for an organization to meet all legal requirements and still act in an unethical manner. Forum participants concurred that organizations must weigh both legal and ethical considerations in their decision-making processes.

While AI Ethics are not binding, associated principles and values have been and continue to be codified into laws, which are binding. As ethical standards are incorporated into laws, institutions continue to face situations where decisions were made prior to related laws being passed. For instance, the European Union's General Data Protection Regulation (GDPR) came into effect in May 2018, however, European institutions had to make customer data related decisions prior to 2018. This highlights the need to continually refine and update practices based on evolving standards and laws.

The following pyramid hierarchy illustrates the importance of considering both legal and ethical aspects while balancing them with the organization's values (see Figure 2). At the bottom tier of the pyramid is "Organizational Values" — decisions need to be made to support values that the organization deem important. The next level is "Ethics" — aspects of Canadian values that are generally accepted but not explicitly codified into laws. The top of the pyramid is "Law" , as all institutions looking to do business in Canada need to abide by laws.

Figure 2
Different levels of ethical decision-making



Challenges and Considerations

The subjectivity of ethics, the recognition that ethical standards change over time and their codification into laws and regulations show the challenges and complexity of addressing AI Ethics. For financial institutions operating in multiple jurisdictions the challenges and complexity are compounded.

A number of considerations to mitigate those challenges were discussed at the forum:

- **Multidisciplinary Views.** Broad perspectives through the engagement of multidisciplinary and diverse teams (e.g., computer scientists, lawyers, financial data scientists, ethicists) are needed at all stages of development and use of AI applications.

- **New Roles.** Organizations could consider greater corporate investment in AI ethics by adding new roles such as a Chief Ethics Officer or Chief Trust Officer.

- **Standards.** Standards are distinct from regulations or laws; however, they may be voluntary or mandatory. Standards setting bodies could set agreed upon ethical guidelines for the financial services industry to help manage the related risks of AI technologies. Professional associations could also develop ethical standards. The Canadian Institute of Actuaries (CIA) have developed minimum ethical standards and their members are expected to comply with the ethical framework.

- **AI Designation.** To ensure that appropriate ethical questions are considered during the process of designing, developing, and implementing AI systems, it was suggested that a designation be created requiring training in ethical issues. For example, Singapore has created a Chartered Engineer for AI designation that addresses this.

Business Goals, Data Bias and Unfairness

Forum participants discussed various approaches to address issues of bias, particularly related to historical and ongoing processes that have left some groups disadvantaged. Discussions also included the role of financial institutions in creating a more equitable and fairer world and in bringing about social justice.

Data used for AI training and development can be the source of bias and unfair outcomes. One approach to address potential discriminatory bias is dubbed ‘fairness through unawareness’, where financial institutions would not use certain personal attributes in model development. This would render the models ‘attribute-blind’ but may not be outcome neutral.

Further, in some cases, excluding protected or sensitive variables from the AI training process does not necessarily lead to ‘unaware’ outcomes. Retained variables could act as proxies for the excluded variables and lead to unfair outcomes. A dilemma could then arise if excluding protected and proxy variables leads to a significant reduction in available data and potentially poorer performance of AI models. This would need to be resolved based on the financial institution’s ethical values.

Another approach would have financial institutions investigate and ensure fairness by testing models against various personal attributes. Depending on the results of the model testing, financial institutions might modify models or create separate models for sub-populations of customers to ensure uniform treatment. However, this would require collecting and recording those attributes. Exclusion of protected variables has been promoted for fairness and privacy purposes and collection of protected variables is not permitted in a number of jurisdictions. However, absence of those variables precludes the ability to assess AI fairness. Forum participants agreed that a change in culture is necessary to see the collection and use of information as helpful for evaluating modeling decisions, instead of being disadvantageous to consumer groups. The collection and use of protected variables for AI is an area for further exploration.

Nonetheless, it could be a good practice if organizations vet their data strategies with legal, compliance, and marketing teams in order to ensure that their objectives meet regulatory and legal requirements for data usage and as well as with customers' expectations. This can help organizations to avoid legal and regulatory compliance issues, and to build trust with customers by being transparent about how their data is being used.

It was recognized that societal expectations that financial institutions maintain high ethical standards continues to increase. In addition, there is real reputational risk and associated consequences when harm, actual or perceived, is done to customers.

Fairness

There is no universal definition of fairness. What is perceived as fair depends strongly on the context. Within the realm of algorithmic fairness, there are different mathematical definitions, some conflicting with one another.[14] When fairness is raised within the AI context, it tends to be the avoidance of discrimination against persons or group of persons. From a legal perspective, the Canadian Charter of Rights and Freedom presents clear rules for avoiding discrimination.

Unfairness can be against an individual or a particular group, as such 'individual' and 'group' fairness measures have been developed. Individual fairness would mean parity between similar individuals while group fairness would mean parity between groups, such as demographic groups. While both are important, they are different definitions, and it is generally not possible to optimize both measures at the same time.

Despite the legal perspective, complying with the law does not always mean that actions and outcomes are fair or perceived to be fair. For example, marketing products to only certain groups of people can be seen as unfair by some people and fair by others. Such targeted marketing by financial institutions arises from AI generated segmentation in order to deliver the right message or product to the right customer in a cost-effective way.

Much rigor would need to be undertaken to understand fairness in decisions made or actions taken based on AI generated results. In addition, financial institutions would need to define what is considered "fair" depending on the particular use of an AI application.

Bias

Bias is a term often used when unfairness is discussed, and they are often used in the same context. Bias is commonly defined as "*inclination or prejudice for or against one person or group, especially in a way considered to be unfair*."

Bias can arise from different sources and can take varying forms. For example,

- bias resulting from sensitive or protected variables, such as gender, religion, or ethnicity, can be due to historical social factors reflected in the data, lack of representation in data collected by organizations or improper data collection.
- data scientists can introduce bias through the choice of variables or algorithm selected.
- humans who use AI model results to make decisions could introduce bias by overriding AI model outcomes.
- bias can appear in an improperly trained model even when the training data is not biased.
- bias can arise from the choice of outcome measure.

However, bias also has a more technical meaning, aligned with model objectives, and this definition of bias is the desired outcome of a model. For example, insurance models show a bias against drivers with poor driving records and causes these drivers to pay higher premiums.

Assessment and Intervention

Outcome or fairness metrics are ways to detect discrimination, so it was recommended that institutions have governance frameworks similar to those used to monitor model performance. While fairness could be considered by the financial institution through new (ex-ante) model assessments, reviews, and designs; a gap could exist for legacy models that were not subject to these reviews.

It was also recognized that outcome measurement for fairness is context dependent as such different fairness measures would apply in different contexts. For example, fairness in a credit granting decision may be defined differently than in a hiring process. As a result, different fairness measures would be more appropriate to use in different contexts.

Though bias and unfairness could arise from data, it also holds some solutions. Financial institutions could emphasize data representation, explore the use of synthetic data, or use in-processing or post-processing techniques to address discrimination.[15] As has been highlighted above, protected variables would normally be needed for assessment and intervention.

Adversarial ML in Making Fair Decisions

Harrison Edwards and Amos Storkey, University of Edinburgh, used Adversarial machine learning (ML) to address the problem of fairness in ML. They defined a decision as fair if it does not depend upon sensitive variables such as gender, age, or race. Their approach was to construct synthetic data that preserves information about the original data except for the dependency on the sensitive variable.[16]

Privacy and Right to Recourse

The financial services industry is subject to greater scrutiny than other industries with regards to customer consent and privacy. Organizations are expected to devote adequate resources to ensuring privacy and protecting customers' data.

With the increasing use of AI driven decisions, there is a need to ensure that customers are provided appropriate disclosure and transparency on how their data is used, and that they have avenues of recourse for AI decisions that are efficient and effective.

Customer Consent

Subjectivity of ethics and values could impact consent. For the same data, some people may freely provide consent while some people would have concerns. Customer consent is required when collecting and using their data. In this respect, forum participants identified challenges and came up with certain recommendations. During the forum discussions, participants coined a term "consent drift" that refers to the case where customers provide consent for data to be used for a particular purpose, however over time the same data is used for a different one. Such case would necessitate ongoing consent management.

To prevent negative effects on customers, financial institutions should make sure that customers give serious consideration to what their consent implies. This could be achieved through making the consent documents (e.g., Terms and Conditions) that are short, easy to read and understand. It is critical for financial institutions to account for challenges that customers may face when providing their consent. For instance, a single piece of data may seem harmless on its own, but when it is combined with other data, it could have unintended implications. It can be difficult for customers to anticipate these potential consequences when providing consent. While consent is necessary and should be requested, it was noted that the inability to obtain consent from customers or potential customers to use their information may hinder the ability of financial institutions to tailor products to existing customers or foster financial inclusion.

Digitization was identified as one of the factors that can exacerbate customers' awareness of consent implications. The use of digital forms could make it harder for customers to ask questions, seek clarification, or have a back-and-forth discussion with the organization's employees.

Certain ethical and (possibly legal) issues could emerge from information collected at scale when consent is impossible to obtain. For instance, it could result in privacy violations which can be particularly concerning when sensitive personal information, such as financial and medical data, is collected.

Achieving Privacy of Data

Privacy enhancing techniques enable collaboration and the sharing of sensitive information in a privacy preserving manner. The techniques include homomorphic encryption, federated learning, secure multiparty computation, differential privacy, and pseudonymization, among others.

Note that using some legacy techniques, such as pseudonymization, does not guarantee complete privacy of data, and organizations should take that into consideration. For example, certain types of private information can be reidentified when pseudonymized data sets are combined with other data sets.

While these techniques have shown much potential in some use cases and hold much promise for the future, their general use does not inherently mean compliance with privacy laws such as the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA).

There could also be non-technical ways, such as appropriate governance mechanisms, that can be helpful in improving data privacy and to some extent ensuring protection for data and models.

Operationalizing AI Ethics and Organizational Structures

Operationalizing AI ethics is critical. Organizations should maintain transparency, both internally and externally, through disclosure on how they ensure high ethical standards for their AI models. Furthermore, since ethical standards change, it is necessary to document the rationale for decisions made. Documentation is also necessary for auditing purposes.

While it was recommended that third parties or independent bodies be used to conduct assessments for risk impact, privacy, bias, and fairness, this presents significant accountability challenges. Some of those are outlined within the 'Explainability' chapter of the "EDGE" principles, under Third-party disclosures. Furthermore, such an independent body would need to have access to the protected variables in order to conduct assessments. The use of third parties could be an area for further exploration.

Regulations

Regulations support society by ensuring the safety and soundness of the financial system and protecting consumers.

There has been an ongoing discussion about striking the right balance between regulations and innovation, that is, setting robust regulations while ensuring financial institutions continue to innovate and remain competitive.

AI has and will continue to provide benefits to financial institutions and to their customers. With such benefits, it is expected that financial institutions will increasingly embed AI within their products, processes, and decision-making. However, the realization of the new risks and exacerbated risks such technology could pose have necessitated various jurisdictions to begin to formulate regulations.

The forum addressed the following key questions on regulation:

- What is the state of AI regulations globally?
- What is expected from financial institutions?
- What are the positions of regulators?

> **"AI is a tool. How you use and regulate it is context specific."**
>
> Oliver Carew; Senior Manager Fintech at EY, previously with The Bank of England

State of AI Regulations and Policies across Jurisdictions

With increased awareness and discussions on the benefits and the risks with AI, regulators and policymakers are taking action to ensure those benefits continue to be realized while the risks are prevented or mitigated. This is important to contribute to the public confidence in the use of AI and in financial systems.

In recent years, policy makers and regulators in various jurisdictions have been reviewing existing laws and regulations, seeking input from stakeholders regarding AI, and drafting new regulations and policies to address AI risks.

Forum participants explored the recent activities being carried out across jurisdictions and associated publications, within the European Union (EU), the United Kingdom (UK), the United States (US) and here in Canada.[17] Most of the publications are in draft form, however, they give indication of regulations and policies that financial institutions will need to comply with.

Approaches to AI principles and regulations from the various jurisdictions are varied in terms of scope and impact, however, they align with respect to:

- AI models should be conceptually sound (e.g., accurate, reliable, robust, sustainable).
- Organizations should have proper governance structures that address challenges created by AI (e.g., transparency, accountability).
- Explainability is essential in high-stake decisions (including those with customer impact).
- AI should not cause any harm to individuals and society (e.g., bias, discrimination, ethical considerations, privacy concerns).

Characteristics of Successful Regulations

Forum participants explored the quality and content of regulations that should make them successful for the financial industry, that is, fostering innovation and addressing associated risks.

With increased development and adoption of AI, a principle-based approach was seen as the best way to ensure guidance remains robust over time. In addition, clear and balanced regulations were deemed important. For instance, institutions want to have the ability to innovate; they would prefer to have regulations that are clear regarding what is allowed, yet not so prescriptive that they stifle innovation.

Consistency

Financial institutions will need to comply with expectations from different regulators. In addition, some institutions operate in different jurisdictions. The need for consistency was raised in order to prevent regulatory burden and cost of compliance. While consistency across different countries would be challenging, it was important to ensure consistency within Canada, that is, among Federal and Provincial regulators. The need for alignment between regulations and laws was also highlighted considering the published draft Bill C-27, which covers privacy, data, and AI. Regulations should be adapted to a Canadian context to ensure it meets the populations' needs and is culturally acceptable.

Best Practices

Principle-based guidance often leaves room for interpretation. Forum participants indicated that further guidance on best practices for effective AI implementation and governance is needed. These best practices would be industry recommendations and not form regulatory expectations.

Third Party

The use of third-party data and AI products is inevitable, more so for smaller institutions. Standards around third-party risk management and/or independent review of the third parties were deemed necessary, with consideration given to avoid regulatory arbitrage between internal and vendor-provided solutions. Such standards can also cover the types of third-party data that can be used by financial institutions. When data is obtained from third-parties, certain guardrails are essential to establish for data provenance, data lineage, and data quality.

Feedback
Stakeholder input was deemed necessary while developing guidance. In addition, there should be a mechanism to receive feedback from stakeholders to allow for enhancement of regulations.

Sandbox and Platforms
Creation of regulatory sandboxes could also allow institutions to innovate and test new ideas with awareness and involvement by the regulators. It was also suggested that regulators encourage the industry to create a platform where financial institutions can discuss key topics related to AI and explore best practices.

Proportionality
Regulations should account for differences in size, materiality, and organizational capabilities across financial institutions. Smaller financial institutions might need to rely on external parties to as they adopt AI.

In addition to the quality and content for successful regulations, forum participants recommended that regulatory bodies promote AI literacy, including data and consent, and also encourage financial institutions to do the same in order to broaden financial inclusion in Canada.

Voice of the Regulators

While it was beneficial to seek perspectives from the forum participants regarding development of sound and successful regulations, it was also important to get perspectives directly from regulators. Despite the financial sector being ahead of many other sectors with regards to understanding the risks from models, it is necessary for regulatory bodies to keep abreast of the new risks emerging from adoption of AI.

Harmonization
While regulators have different mandates and scope, there is a need to harmonize regulations as best as possible. Consideration should also be given to issues that could arise as new technologies or solutions, such as open banking, converge with existing technologies. While there is a need for alignment between Canadian regulators, there is also need for alignment between international jurisdictions.

Innovation
While financial institutions need to innovate and manage the risks that arise from this, regulators also have to innovate, test and learn, as well as glean insights from AI use in other industries. Some regulators already

have an innovation office. It is important that regulators not be perceived as a hindrance to financial institution innovation. Discussing and sharing with the private sector could be beneficial.

Collaboration

Collaboration among regulators can help with regulation harmonization as well as support regulatory innovation. Provincial regulators are smaller and sometimes have a dual mandate, prudential and conduct and sometimes rely on larger Federal regulators. These characteristics make collaboration important.

Multidisciplinary Focus

Various risks arise or increase with AI, such as legal and compliance. AI needs a different approach for model risk governance. Consideration should be given to AI embedded into software applications and systems. Diversity of thought is fundamental.

Education

With the pace of AI innovation and the impact, industry associations can help educate members on key aspects of AI such as bias, ethics, and model governance as well as the challenges and considerations.

Smaller Financial Institutions

Proportionality and outsourcing need to be considered in regulating smaller financial institutions. Such financial institutions may find it challenging to adopt AI and this may impact their ability to innovate and grow. In addition, those financial institutions may rely more on third parties and/or outsource AI development.

Conclusions

When OSFI and GRI started on the path to organize the FIFAI, there was a vision of bringing industry stakeholders together to move forward on best practices related to some aspects of managing the risks from AI use in the Canadian financial services industry.

As the forum unfolded, it was rewarding to see the engagement and active participation of the attendees, demonstrating the need for opportunities to discuss topics of interest and collaborate in finding possible solutions to challenges. The collaboration between AI experts from different institutions was very enriching and led to important insights.

It was evident that different topics are at different levels of maturity in terms of the understanding of the challenges and the practical solutions that have been found to address them. Out of the EDGE principles, Data and Governance have been part of institutions' frameworks for a long time, while Explainability and Ethics have more recently come up to the forefront. This is also reflected in the length of discussion or level of conclusion for the different topics that was arrived at the forum.

As AI use at financial institutions continues to advance and evolve, forum discussions concluded that stronger trust needs to be built in the technology to help accelerate adoption. Appropriate levels of explainability and disclosure for each use case as well as a customer centric approach that upholds ethical values and protects privacy are needed to promote the trust of customers. A strong governance framework, including data governance, that follows a risk-based approach and is based on multidisciplinary collaboration promotes trust within financial institutions. A harmonized regulation framework that respects Canadian cultural values and is simple to comply with while providing guardrails for innovation promotes trust across the broader financial system and community.

General education on AI, its benefits and limitations are another aspect that was discussed as an enabler for AI adoption. All levels of a financial institution need to be versed in the implications of using AI. A few to consider include: aspects related to data use, data governance and data privacy as well as ethical considerations and customer impact, along with strengths and drawbacks of particular AI techniques. Effectively integrating AI tools into the financial services industry requires a combination of technical expertise, organizational changes, and cultural shifts.

Third-party services were another key point in the discussion, as they emerge in almost every aspect of the development of AI solutions. Tools for data management, data governance, model development, model governance, reporting, model risk management, open-source code are just some of the many possible spaces where third-party services can be helpful in AI adoption. Third-party services that provide help in compliance with regulation or that certify institutions based on industry standards is another area that is emerging and was discussed in the forum. As the field moves forward, there are increasing opportunities to leverage the work of different players, and this also brings in challenges that were not as pervasive before. This field will require further work and exploration as the financial services industry moves forward in the use of AI.

There is still much work to be done. The insights and discussion from the forum are a steppingstone in the way forward for Canadian Financial Institutions' adoption of AI. The forum was a testament to the need for collaboration and a multidisciplinary approach at different levels and the advantages it can bring.

Acknowledgements

We want to express our sincere gratitude to the speakers and participants at the workshops for their meaningful presentations and active engagement which proved consequential to the discussions and the completion of this report.

We would also like to thank the OSFI and GRI teams for their support in this initiative, with special recognition to Romana Mizdrak and Bruce Choy. Our appreciation also extends to Alexey Rubtsov for his coordination and Obim Okongwu for his management of the initiative. We want to express our gratitude to the OSFI Procurement, Legal and Communications teams, the design and facilitation teams, the report writing team, and the GRI Communications team. Finally, thanks to Rotman School of Management for providing space.

Speakers

Future of AI

- Stuart Davis; EVP, Financial Crimes Risk Management & Group Chief Anti-Money Laundering Officer at Scotiabank
- Foteini Agrafioti; Chief Science Officer at RBC and Head of Borealis AI
- Andrew Moore; Vice President and General Manager, Cloud AI and Industry Solutions at Google

Explainability

- Alex Wong; Professor, Canada Research Chair in AI and Medical Imaging at University of Waterloo
- David Heike; Managing Director, Head of Risk Modeling — Consumer & Community Banking at JPMorgan Chase & Co.
- Agus Sudjianto; EVP and Head of Model Risk at Wells Fargo

Data

- Ima Okonny; Chief Data Officer at Employment and Social Development Canada

Governance

- Donna Bales; Co-Founder of Canadian RegTech Association
- David Palmer; Senior Supervisory Financial Analyst at Federal Reserve Board, United States

Ethics

- Carole Piovesan; Managing Partner at INQ Law

Perspectives from Other Jurisdictions

- Oliver Carew; AI expert & Senior Manager Fintech at EY, previously with The Bank of England
- Qiang Zhang; Deputy Director, AI Development Office at Monetary Authority of Singapore (via Interview)

Speakers

Perspectives on AI

- Bruce Choy; Managing Director, Research, Global Risk Institute
- Romana Mizdrak; Managing Director, Risk Quantification, OSFI

Perspectives from Canada

- Alexey Rubtsov; Associate Professor, Dept. of Mathematics, Toronto Metropolitan University
- Ana Garcia; Director, Risk Quantification, OSFI

Showcase Presentations

- Benoit Deshaies; Director — Data and Artificial Intelligence, Treasury Board of Canada Secretariat
- Bradley Fedosoff; SVP Architecture, Data & Analytics, CIBC
- Eugene Wen; VP Group Advanced Analytics, Manulife
- Greg Kirczenow; Senior Director — AI Model Risk Management, RBC
- Michaela Capra; AVP — Corporate Risk Digital Innovation, Sunlife
- Shingai Manjengwa; Director — Technical Education, Vector Institute
- Stephanie Kelley; Assistant Professor, Ivey Business School at the University of Western Ontario

Participants

- **Amex Bank of Canada**, Pat Smith
- **Antara Risk Management**, Sanjiv Talwar
- **Bank of Canada**, Maryam Haghighi
- **BC Financial Services Authority**, Steven Wright
- **BMO**, Drew Galow, Letitia Golubitsky, Suyi Chen
- **Business Development Bank of Canada**, Sherrilyn Lequin
- **Canada Deposit Insurance Corporation**, Neville Arjani
- **Canada Pension Plan Investment Board**, Brendon Freeman
- **Canadian Institute of Actuaries**, Joel Li
- **Canadian RegTech Association**, Paul Childeross
- **CIBC**, Brad Fedosoff, Ozge Yeloglu
- **Decca** and **McKinsey**, Matthew Killi (Formerly with)
- **Financial Services Regulatory Authority of Ontario**, Ivy Ou
- **Financial Transactions and Reports Analysis Centre of Canada**, Nathalie Martineau
- **Global Risk Institute**, Mike Stramaglia, Alexey Rubtsov, Bruce Choy, Mark Engel
- **Innovation, Science and Economic Development Canada**, Anastasiia Tryputen, Surdas Mohit
- **Intact Financial Corporation**, Sebastien Bernard
- **Manulife**, Eugene Wen, Henry Li
- **Northbridge Financial Corporation**, Cheston Chiu
- **OMERS**, Richard Slessor, Sami Ahmed
- **Ontario Securities Commission**, Levin Karg
- **OSFI**, Ana Garcia, Greg Caldwell, Karyn Leung, Mohamad Al-Bustami, Obim Okongwu, Patrick Cane, Regis Dahany, Romana Mizdrak, Sharon Chambers Creary, Stephen Manly
- **RiverRun Ventures GP**, Daniel Moore
- **Royal Bank of Canada**, Greg Kirczenow, Jun Yuan, Dominique Payette
- **Schwartz Reisman Institute for Technology & Society**, Monique Crichlow
- **Scotiabank**, Carrie Chai, Gail Towne, Michel Valentik
- **Sunlife**, Mihaela Capra
- **TD Bank**, Paige Dickie, Baiju Devani
- **Treasury Board of Canada Secretariat**, Benoit Deshaies
- **University of Toronto**, Zissis Poulos, John Hull
- **Vector Institute**, Andres Rojas, Shingai Manjengwa
- **Western University**, Cristian Bravo

References

1 Slack, D., Hilgard, S., Jia, E., Singh, S., Lakkaraju, H.: Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods (2019) (https://arxiv.org/pdf/1911.02508)

2 Rudin, Cynthia: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 206–215 (2019)

3 Notional amount can be taken to be the initial notional amount of the transaction, or the adjusted notional amount based on FX conversion, accrued interest, or stock multiplier.

4 Indeed, the treatment of protected attributes varies: Singapore allows for the collection and use of gender data in AI models; the European Union allows for the collection of gender but prohibits the use of gender as a feature in the training and screening models used for individual lending decisions; the United States prohibits the collection and use of gender data; Canada does not explicitly prohibit the collection and use of protected attributes established under the Charter of Rights and Freedoms for assessing bias. At the federal level, protected attributes are specified under the Canadian Charter of Rights and Freedoms. Additional attributes are specified provincially with variation among provinces. Even within one province, some industries (e.g., insurance) are able to treat customers differently based on protected attributes (e.g., young males receive higher auto insurance rates).

5 The Bank of England's final report on "Artificial Intelligence Public-Private Forum" also provided a similar example of "food labelling", see p.20. (https://www.bankofengland.co.uk/-/media/boe/files/fintech/ai-public-private-forum-final-report.pdf)

6 Canadian Audit and Accountability Foundation. Practice Guide to Auditing Oversight. (https://www.caaf-fcar.ca/en/oversight-concepts-and-context/what-is-oversight-and-how-does-it-relate-to-governance/what-is-governance)

7 The Federal Reserve and the OCC's SR 11-7: Guidance on Model Risk Management (https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm)

8 OSFI's E-23: Enterprise-Wide Model Risk Management for Deposit-Taking Institutions (https://www.osfi-bsif.gc.ca/Eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/e23.aspx)

9 Treasury Board Secretariat, Government of Canada. Directive on Automated Decision-Making (https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592)

10 The Institute of International Finance's Machine Learning Governance Summary Report (https://www.iif.com/portals/0/Files/content/Innovation/12_4_2020_mlg_summaryreport.pdf)

11 Tudor, Deniz; Model Risk Management Lessons Learned: Tracing Issues from the Pandemic to the Great Recession. Global Association of Risk Professionals. July 22, 2022 (https://www.garp.org/risk-intelligence/operational/model-risk-management-lessons-220708)

12 MLOps refers to the practices, processes, and tools that organizations use to manage the production and deployment of models. It is an extension of DevOps, which focuses on automating and streamlining the software development process, to the realm of machine learning. DevOps focuses on software development and deployment, whereas MLOps also includes data engineering and model deployment.

13 Leslie, D.: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. (2019) (https://doi.org/10.5281/zenodo.3240529)

14 See, for example, FEAT Principles Assessment Case Studies (Veritas Document 4), Canadian Charter of Rights and Freedom (https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-ccdl/pdf/charter-poster.pdf), and Towards the Right Kind of Fairness in AI (GETD | AI Research & Thought Leadership, May 2021), among others.

15 Data-driven approaches to reduce discrimination include, down-sampling, up-sampling, gender-aware hyperparameter tuning, probabilistic gender proxy modeling.

16 Edwards, H. and Storkey, A.: Censoring Representations with an Adversary (2016) (https://arxiv.org/abs/1511.05897)

17 The EU AI Act (2021), The UK Report of the AI Private-Public Forum (2022), The UK Artificial Intelligence and Machine Learning Discussion Paper (2022), The UK Model Risk Management Principles for Banks (2022), Singapore Report by Veritas Consortium (2022), The US Request for Information on AI (2021), The US Algorithmic Accountability Act (2022), The US American Data Privacy and Protection Act (2022), Canada's OSFI's Technology Risk Discussion paper (2020), Canada's E-23 Guidelines Industry letter (2022), Canada's Draft Bill C-27 (2022)

Hexagon image designed by kjpargeter / Freepik