



# TC NOTES

PRACTICAL **LEADERSHIP**  
AND **GUIDANCE** FROM  
**TORONTO CENTRE**

## SUPERVISORY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

JULY 2022



# SUPERVISORY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

## Contents

Introduction.....	3
Artificial intelligence and machine learning.....	3
Definitions.....	3
Drivers.....	5
Uses of AI and ML in the financial sector.....	5
Customer service .....	6
Credit scoring and pricing .....	7
Insurance underwriting and pricing.....	8
Trade finance .....	9
Investment .....	9
Trading.....	9
Robo Advice .....	9
Fraud prevention .....	10
Money laundering detection .....	10
Cyber security .....	10
Implications for supervision: risks to supervisory objectives .....	11
Risks to the safety and soundness of financial institutions .....	11
Risks to financial stability.....	12
Risks to consumers .....	13
Regulation and supervision: way forward .....	14
High-level principles for ‘trustworthy’ AI.....	15
Model requirements .....	18
Examples .....	20
Suptech.....	22
Conclusion .....	24
References .....	25

Copyright © Toronto Centre. All rights reserved.

Toronto Centre permits you to download, print, and use the content of this TC Note provided that: (i) such usage is not for any commercial purpose; (ii) you do not modify the content of this material; and (iii) you clearly and directly cite the content as belonging to Toronto Centre.

Except as provided above, the contents of this TC Note may not be transmitted, transcribed, reproduced, stored or translated into any other form without the prior written permission of Toronto Centre.

The information in this TC Note has been summarized and should not be regarded as complete or accurate in every detail.

# SUPERVISORY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

## Introduction<sup>1</sup>

Financial institutions across all sectors are increasingly making use of artificial intelligence (AI) and machine learning (ML). These applications are more widespread in the financial sectors of developed economies, but they are also beginning to appear in emerging market and developing economies. Examples - from both developed and emerging economies - include the credit scoring of loan applications, “chatbots” for communication with customers, and identifying suspicious financial transactions.

The use of AI and ML can bring many benefits and opportunities, including expanded access to financial products and services. But the use of AI and ML may also change the nature of familiar risks (for example, credit and insurance underwriting risks); generate new risks (for example, the “black box” nature of ML applications); and give rise to new considerations around the ethical use of AI and ML (for example, to avoid unfair bias and discrimination).<sup>2</sup>

Where such applications are used it is therefore important for supervisors to understand the risks to the safety and soundness of financial institutions, financial stability, and consumer protection.

This Toronto Centre Note describes some of the uses of AI and ML by financial institutions; considers the supervisory responses to such uses; and highlights some ways in which supervisory authorities can themselves use AI and ML (a subset of “Suptech”) to improve the effectiveness and efficiency of supervision.

## Artificial intelligence and machine learning

### Definitions

The Financial Stability Board (2017) defines **artificial intelligence** (AI) as the “theory and development of computer systems able to perform tasks that traditionally have required human intelligence”.

The use of technology can enable a computer to simulate human behaviour and to perform various complex tasks. This can automate the same routine tasks a human would typically perform, enabling these tasks to be performed more quickly, cheaper, and without human errors. Technology can also be used to automate tasks that humans could conceivably perform, but it would not be cost-effective for them to do so (for example, running multiple calculations using large data sets).

**Machine learning** (ML) is a subset of AI, defined by the Financial Stability Board as a “method of designing a sequence of actions to solve a problem, known as algorithms, which optimise automatically through experience and with limited or no human intervention”.

---

<sup>1</sup> This Note was prepared by Clive Briault. The author is grateful to John P O’Keefe for his insightful comments on an earlier version of this Note. Please address any questions about this Note to [publications@torontocentre.org](mailto:publications@torontocentre.org)

<sup>2</sup> General reviews of the applications, opportunities and risks of AI and ML can be found in Financial Stability Board (2017), International Organisation of Securities Commissions (2021), and OECD (2021).

ML can process large quantities of data using algorithms to generate outputs such as identifying patterns and making predictions.<sup>3</sup> These data can be drawn from increasingly diverse and innovative sources, including personal and corporate transaction details, market data, mobile communications, social media, and digital documents. The techniques used can include data analytics, data mining, and natural language processing. The algorithms can change over time to improve the outputs.

### **Models and algorithms**

A model is a simplified representation of a process that focuses on the key features that are of interest to the modeler. For example, a bank might want to predict the likelihood of a borrower defaulting on a loan, or an insurer might want to predict the likelihood of a policyholder making a claim.

This can be done by relating observations on the outcome (a borrower default or non-default, or a policyholder claim or non-claim) over some interval of time to variables thought to influence the outcome. For example, a loan default might be thought to depend on variables such as the borrower's income, employment and age. Model parameters, such as weights on model variables, determine how each variable influences the modelled outcome. The values for model parameters can be estimated using a variety of algorithms that minimize predictive error based on a 'training' dataset of model inputs (for example, data on loan defaults/non-defaults and variables that influence default).

The ability of the algorithm to model the relationships between the outcome (the dependent variable, such as loan default or insurance claims) and the explanatory variables varies with the algorithm used for model estimation. For example, a linear regression assumes a linear relationship between the explanatory variables and the outcomes. Alternatively, artificial neural networks can model virtually any functional relationship between model variables and outcomes, including complex and non-linear relationships.

In the final step, the estimated model can be applied to new data on explanatory variables to make predictions of the outcome.

ML can involve different types of 'learning'. The Financial Stability Board (2017) distinguishes between the following:

*Supervised learning* – where the algorithm is fed a set of 'training' data that contain labels on some portion of the observations (for example, some data may be labelled as suspicious, and some data may be labelled as clean). The algorithm can then 'learn' a general rule of classification and use this to predict the labels for the remaining observations in the data set.

*Unsupervised learning* – where the data provided to the algorithm do not contain labels. The algorithm is asked to detect patterns in the data by identifying clusters of observations that depend on similar underlying characteristics.

*Reinforcement learning* – where the algorithm is fed an unlabelled set of data, chooses an action for each data point, and then receives feedback (perhaps from a human) that helps the algorithm learn.

*Deep learning* – where algorithms such as artificial neural networks work in 'layers' inspired by the structure and function of the brain. This can be used to analyze unstructured data

---

<sup>3</sup> It is important to note here that the patterns identified by ML are correlations with other events or patterns – ML does not determine causality.

such as images, audio recordings, or texts. Deep learning can be used for supervised, unsupervised, or reinforcement learning.

## Drivers

Drivers of the increasing use of AI and ML by financial institutions include:

- Advances in technology (in particular computing power and analytic techniques).
- The massive expansion in the availability of data and information, not least from digital sources.
- The rapidly reducing cost of data storage (in particular through cloud computing) and data analysis.
- The greater willingness and ability of (some) households and corporates to use digital channels, and the impact of the COVID-19 pandemic on the increasing use of digital channels for financial transactions.
- The scope to reduce costs and increase (or at least maintain) profitability in an increasingly competitive marketplace (including competition from new entrants).
- The greater predictive accuracy that AI and ML models may generate.<sup>4</sup>

It is therefore becoming increasingly commercially viable for financial institutions to use AI and ML to support their products and services. However, this requires high levels of investment, the availability of large data sets, access to research, and human capital, all of which may be less readily available in many emerging market and developing economies.<sup>5</sup> This may explain why AI and ML are currently only being used in a small number of emerging market and developing economies.<sup>6</sup>

## Uses of AI and ML in the financial sector

Many uses of AI and ML have emerged in the financial sector. Some examples are described in this section. All of these examples offer some potential benefits and opportunities from the use of AI and ML, including:

- Improved customer experience.
- New and improved financial products and services.
- Lower costs for financial institutions and a more efficient financial services sector.
- Improved compliance and risk management.
- More effective identification of fraud and money laundering.
- Greater financial inclusion.

The use of AI and ML could also bring significant benefits to emerging market and developing economies, enhancing access to credit and insurance; identifying customers and documents to support account opening; improving customer service; supporting compliance

---

<sup>4</sup> Many research papers show that AI and ML models can be more accurate than standard linear regression models. However, AI and ML models may not be so useful when the variables (explanatory variables and the dependent variable) fall outside the ranges used in the 'training model' (see O'Keefe 2022), and – as with standard models – AI and ML models will perform less well if some of the explanatory variables are highly correlated, or if the data contain extreme outliers.

<sup>5</sup> There may also be less scope to exploit economies of scale in emerging market and developing economies – the fixed cost of establishing an AI and ML capability may only generate sufficient benefits if it can be applied to multiple applications and large data sets.

<sup>6</sup> International Monetary Fund (2021) discusses a broader concern that new technologies may widen the gap between rich and poor countries by shifting more investment to advanced economies where the data and technology are already more established.

with anti-money laundering requirements; and detecting and preventing fraud and cyber-security threats.<sup>7</sup>

## Customer service

AI and ML have supported the development of chatbots and virtual assistants to improve interactions between humans and machines. Natural language processing engines<sup>8</sup> can provide increasingly realistic and useful interactions, learning from previous interactions, and adapting to different types of customers and different customer behaviours.

The simplest use of these applications is to answer customer questions, providing customer access 24 hours every day without the customer having to visit a branch or wait to speak on the phone with a customer service representative. This can be used for basic account queries, questions about services and products, basic insurance claims handling, and making complaints.

Applications can also be more proactive, allowing financial institutions to contact customers to offer them products and services and to make recommendations on budgets, spending and saving, and insurance options. These more customized and personalized contacts can be generated through customer profiling based on big data analytics, using both customer financial transaction history and other data such as customer search engine activity and social media. Such profiling can also be used to determine the best channels through which to contact customers proactively, when to initiate these contacts, and which products and services these contacts should focus on.

A further use of natural language processing is for financial institutions to monitor sales calls, and to check promotional materials and contracts, to test compliance with internal processes and procedures, and with legal and regulatory requirements.

World Bank (2021) describes some examples of the use of chatbots:

Bank BCP (Peru) - a personalized chatbot, Arturito, assists customers in converting currencies, meeting credit card repayments, and accessing 24-hour customer support via Facebook.<sup>9</sup>

Banco Bradesco (Brazil) – a chatbot is available to answer questions relating to 62 products.

MTN (Ivory Coast) – incorporates AI support into its digital wallet, MoMo, to enable customers to better understand their financial products and obligations.

---

<sup>7</sup> World Bank (2021) discusses various ways in which AI and ML could benefit emerging and developing economies, and provides examples of early applications in these economies. It also discusses the roles of financial institutions, governments and financial supervisors in facilitating and fostering the responsible and effective use of technology in the financial sector.

<sup>8</sup> Natural language processing (NLP) refers to the branch of AI that enables computers to understand text and spoken words in much the same way human beings can. NLP combines rule-based modelling of human language with statistical, machine learning, and deep learning models, to enable computers to process human language in the form of text or voice data and to 'understand' its meaning. NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly.

<sup>9</sup> See <https://cioperu.pe/articulo/23054/bcp-lanza-arturito-un-chatbot-que-funciona-en-facebook-messenger/>

## Credit scoring and pricing

AI can be used simply to automate existing processes for loan applications and monitoring borrower performance, making these processes quicker and cheaper, and to remove human errors.

There is also scope to use AI to undertake more sophisticated analysis, using a wider range of data and information than the traditional reliance on a potential borrower's income, financial transactions with the lender, and credit history (including information available through credit bureaus, where they exist). This can extend the credit analysis to include information that was previously available but was not cost-effective to use.

Beyond this, AI and ML can be used to analyze not only structured data but also unstructured data and information, for example social media (the digital footprint of a customer's behaviour and spending patterns) and search engine activity; mobile phone usage, mobile money transaction data and text messaging activity; and an assessment of non-credit bill payments (for example, the timely payment of mobile phone and other utility bills). This can be used to identify patterns, predict loan quality, spot fraudulent applications, and learn from each interaction.<sup>10</sup>

For lending to companies, AI and ML can be used to analyze company history on digital sales platforms and to analyze social media comments on companies as an input to sentiment analysis.

The analysis of additional data and information can facilitate cheaper, quicker, and potentially more accurate credit scoring and pricing decisions. However, banks and other lenders generally have not completely automated their credit scoring decisions – they typically have 'rule sets' that can override model-based decisions for credit approvals and credit limits; and systems that can trigger reviews of selected decisions by credit analysts.

AI and ML applications can also enhance financial inclusion by facilitating lending to the underserved with no credit history and possibly no account transaction history (because the potential borrower had previously only used cash and informal financial services). Credit decisions can instead be based on alternative data sources and the application of ML algorithms to assess a potential borrower's ability and willingness to repay. These applications can also reduce the cost of reaching previously underserved borrowers by automating processes and making low-value transactions cost-effective.

The International Monetary Fund (2021) cites research showing that AI and ML predictive models can reduce the rate of non-performing loans, benefit underserved loan applicants, and improve the evaluation of small enterprise borrowers, including those owned and run by women who may have previously had no dealings with the formal financial system.

World Bank (2021) describes many examples of the use of AI and ML for credit scoring in emerging market and developing economies, including:

M-Shwari (Kenya) – instead of maintaining a large network of offices and credit agents, the company allows customers in remote and underserved regions to apply for loans through their mobile phone, and then uses AI and ML to review applications and predict the probability of default. The company works in partnership with M-Pesa and NCBA.<sup>11</sup>

<sup>10</sup> See International Monetary Fund (2019) for a more detailed and technical review of the uses of AI and ML in credit scoring.

<sup>11</sup> See <https://ke.ncbagroup.com/m-shwari/>

M-Kajy (Madagascar) – a mobile phone-based microloan facility available to customers using Orange Madagascar for deposits, saving and payments, operated in partnership with First Micro Finance Agency (PAMF).<sup>12</sup>

Branch – a mobile app digital lender operating in Kenya, Nigeria, Tanzania, Mexico, and India, offering microloans to first-time borrowers and customers without bank accounts. Branch uses ML to credit score potential borrowers based on thousands of data points on the individual (the potential borrower has to consent for Branch to access their mobile phone data) and the accumulated experience across borrowers.<sup>13</sup>

MyBucks (Africa) – provided microloans and insurance directly to customers in 12 countries, including Zambia, Malawi, and Uganda, by applying AI technology to analyze data from a potential borrower's mobile phone to generate a lending profile.<sup>14</sup> However, in early 2022 MyBucks was put into bankruptcy by the Luxembourg tax authority.<sup>15</sup>

Aavas (a specialist housing finance company in India) – uses data analytics and AI tools to assess the creditworthiness and willingness of households with informal employment and undocumented incomes to repay loans.<sup>16</sup>

FarmDrive (Kenya) – an agricultural data analytics company delivering financial services to unbanked and underserved smallholder farmers, while helping financial institutions increase their agricultural loan portfolios.<sup>17</sup>

## Insurance underwriting and pricing

The use of AI and ML in insurance underwriting is conceptually very similar to its use in credit scoring – to enable automation, and to use and analyze additional data and information.<sup>18</sup>

Two extensions of particular relevance to insurance are:

- The use of AI to monitor and incentivize behaviours, for example using tracking devices to monitor how carefully a vehicle is being driven, whether security is activated at a property, and whether factories are complying with safety and emissions requirements. Such monitoring can be used to assess the validity of insurance claims, and for the pricing of insurance at renewal.
- The use of automation and ML to assist in the assessment and processing of claims. In particular, high-volume, low-value insurance claims can be dealt with more quickly and efficiently. AI technology can streamline the administrative process through process automation; natural language processing can be used to analyze policy wordings, claim documents, health forms and other documents; and outliers and potentially fraudulent claims can be highlighted.

---

<sup>12</sup> See <https://www.tom.mg/media-actu/pamf-orange-madagascar-m-kajy-instant-loan-via-mobile/?lang=en>

<sup>13</sup> See <https://branch.co/>

<sup>14</sup> See <https://www.forbes.com/sites/mfonobongnsehe/2018/08/02/how-fintech-firm-mybucks-plans-to-offer-access-and-financial-inclusion-to-africas-unbanked/?sh=7d67d0c2f6bc>

<sup>15</sup> See <https://www.itweb.co.za/content/raYAyqodmZ9vJ38N>

<sup>16</sup> See <https://www.aavas.in/>

<sup>17</sup> See <https://techmoran.com/2021/04/13/farmdrive-provides-kenyan-farmers-with-alternative-credit-scoring/>

<sup>18</sup> OECD (2020) describes AI and ML applications in the insurance sector.

## Trade finance

Trade finance transactions are traditionally paper-based, providing scope for the use of AI and ML both to automate the process (reducing the cost of trade financing) and to undertake compliance checks of documentation (extracting data from documents, and highlighting deviations). Trade finance staff can then focus on more complex trade finance transactions and on the anomalies identified by ML.

World Bank (2021) highlights the Tradeteq platform, which simplifies and automates the trade asset distribution process, covering letters of credit, guarantees, supply chain finance and receivables. This facilitates bank-to-bank and bank-to-insurance company transactions, and the purchase of receivables portfolios by institutional investors. The platform also uses alternative sources of data to provide credit analysis.<sup>19</sup>

## Investment

AI and ML can be used to process and analyze large, complex, and diverse data sets (structured and unstructured data, including company reports and other filings, research, and news releases) to identify real-time patterns and investment opportunities. This can be combined with the analysis of market developments, trends, and predictions (for securities, foreign exchange, commodity, and other traded markets) to inform investment decisions and to optimize portfolios.

## Trading

Algorithmic trading refers to the use of ML algorithms to conduct trades autonomously. The algorithms – which can learn and adapt – monitor large volumes of data and information to detect and predict factors that are correlated with changes in market prices, which can then be combined with predetermined sets of instructions (on timing, quantities, and prices) to place a trade on behalf of a trader but without the trader's active involvement.

High frequency trading is a subset of algorithmic trading, using complex algorithms to analyze multiple markets and powerful computer programs to transact a large number of orders in fractions of a second, in part to exploit small price differences.

A closely related use of AI and ML is to enhance the modelling of the impact of a trading firm's trades on market prices, especially for less liquid securities where data on comparable past trades are scarce.

## Robo Advice

Automated financial advice to investors can be cheaper than using human advisers, and can therefore be applied to smaller portfolio sizes, potentially reducing the cost and extending the range of advice to a wider range of investors.

At its simplest, robo advice is an online tool for basic financial planning. An investor enters details of their financial position and personal circumstances, saving and retirement goals, and risk tolerance into the system, and the algorithms then establish a financial portfolio of cash, bonds, equity, life insurance, pension contributions, etc that should be suitable for the investor.

Extensions to the simplest model include generating advice that makes specific investment recommendations (specific products, not just generic portfolio allocations); making investments and undertaking other transactions; and continuous portfolio management

---

<sup>19</sup> See <https://www.tradeteq.com/>

(monitoring and periodically rebalancing a portfolio as market and personal circumstances change).

World Bank (2021) notes that the deployment of robo-advisors in emerging economies is currently limited to those with significant savings pools, such as Brazil, China, and India.

## Fraud prevention

AI and ML can be used to spot and halt (or alert to) potentially fraudulent financial transactions, by scanning large data sets and detecting unusual activities (anomalies) in real-time. Financial transactions can be compared against other data points, such as a customer's account transactions, spending patterns, location, and IP address, so that a transaction (such as a withdrawal, transfer, or card or mobile money purchase) can be automatically declined and flagged for further investigation.

This can reduce fraud and scam activity, protect customers, and potentially boost confidence in digital financial transactions. However, any fraud prevention techniques need to balance the number of declines to avoid either an excessive and unjustified number of 'false positives' (flagged transactions that turn out to be non-fraudulent) or a failure to identify a significant number of fraudulent transactions. ML can potentially optimize this balance by learning from previous experience.

## Money laundering detection

Similar to fraud prevention, AI and ML can be used to detect suspicious activities and transactions for anti-money laundering purposes (for example, unusually large deposits or withdrawals).

AI and ML can also be used to upgrade or transform 'due diligence' identity checks. One stage of this would be to check identity documents against database results from credit bureaus, government agencies, police registers, social media, and other sources to verify identity, and to check whether a government-issued ID is genuine or fraudulent.

A second stage would be to shift from a reliance on physical identity documents to voice recognition, facial recognition, or other similar biometric data.

## Cyber security

AI and ML can be used to automate cyber threat detection, identify compromised data and information, steer investigations after incidents, and extract information to share with other financial institutions and with the relevant authorities.

AI and ML can also be used to transform or upgrade identity checks within a financial institution from a reliance on usernames and passwords to voice recognition, facial recognition, or other similar biometric data, to reduce the risk of compromised or shared usernames and passwords.

## Implications for supervision: risks to supervisory objectives

Why might a financial supervisor worry about the use of AI and ML by financial institutions? This is because alongside the benefits and opportunities there are also some increased risks – to financial institutions, to consumers of financial products and services, and to financial stability.

There are various ‘transmission mechanisms’ whereby AI and ML risks **feed into, and perhaps amplify, familiar areas of concern for supervisors**<sup>20</sup>, including:

**Safety and soundness of financial institutions** – the use of AI and ML may change the quality of credit and insurance underwriting; expose financial institutions to cybersecurity, outsourcing and operational resilience risks; weaken corporate governance; and threaten business model viability.

**Consumer protection** – the use of AI and ML could increase the risks of mis-selling; poor standards of advice; breaches of data privacy and protection; fraud and scams; and financial exclusion.

**Financial stability** – the use of AI and ML could increase systemic risks arising from interconnectedness, concentration, and herd-like behaviours.

There are also some areas where inherent risks from the use of AI and ML pose **new challenges for regulation and supervision**. These include:

**Opacity** – the difficulty of ‘explainability’. Decisions based on AI and ML may be difficult to explain within a financial institution (for example to the board and senior management), to customers affected by the decision, and to supervisors.

**Robustness** – ML algorithms may not be robust to structural shifts and may therefore generate poor results when conditions change. Financial institutions using such applications therefore need to have systems to identify when models are degrading, and to respond accordingly.

**Fairness and ethics** – decisions may be biased and discriminatory, based on gender, age, marital status, race, ethnicity, or geographic location; may use data (or proxies for data) that society judges to be an unacceptable use of data; or may be based on deliberately embedded biases that are intended to harm customers, or on unintentional biases that exclude or harm particular types of customers. Financial institutions can take steps to mitigate these challenges, for example by screening explanatory variables (model inputs) to exclude any that should not be used in their models, and by conducting impact analysis of models to check whether some model inputs are acting as proxies for race, gender and other characteristics that should not be used.

### Risks to the safety and soundness of financial institutions

Financial institutions using AI and ML (be they incumbent institutions adapting their business models, fintech start-ups, or ‘Bigtech’ new entrants) are vulnerable to a range of risks:

The **increasing use of technology, big data, AI and ML**, and the increasing complexity and interconnectedness of financial systems (arising from digitalization, customer access, open banking and other forms of third-party access, and systems interoperability), create vulnerabilities to cyber attacks, fraud, and loss of data. AI and ML applications may also be vulnerable to manipulation intended to change the results (for example, sophisticated

---

<sup>20</sup> There is a parallel here to the ways in which climate-related risks may feed through to ‘traditional’ credit, insurance, market, operational and reputational risks. See Toronto Centre (2021c).

borrowers or traders submitting false information in an attempt to alter the outcomes of AI and ML applications).

The **increasing reliance on third party suppliers**, including for AI and ML, may lead to financial institutions not understanding fully the services being provided by third parties, and not being well placed to monitor and control this outsourcing. While these financial institutions remain responsible for the functions they outsource, this responsibility may prove to be more in name than in substance.

Financial institutions may lack **operational resilience** – they are more exposed to IT failures and to the failure of third-party providers, and at the same time may lack the ability to recover and respond effectively when operational failures do occur.<sup>21</sup>

The **board and senior management may not understand** the uses, complexities and implications of AI and ML within a financial institution (or by its third-party suppliers) and may therefore not put in place the necessary governance, controls, and risk management. This is another aspect of ‘explainability’ and the risk that the use of AI and ML may create ‘black boxes’ in decision-making. The role of human judgement remains critical at all stages of AI and ML deployment, from input of datasets to evaluation of outputs.

**Non-traditional data and information may prove to be of limited or short-lived value**, with patterns and predictions shifting rapidly in the absence of plausible causal relationships. Financial institutions may lack an understanding about the performance of ML across different financial cycles. ML systems that have performed well in a reasonably stable data environment could quickly deteriorate in periods of rapid structural shifts. For example, during the COVID-19 pandemic some ML credit models performed poorly because they did not recognize the role of government support in protecting borrowers from the impact of a loss of income.

Financial institutions may face legal, regulatory, and reputational risks if **some data and information act as proxies for factors that they are prohibited from using**. For example, in some countries, lenders are prohibited from discriminating against borrowers based on race or gender. But ML applications can recreate such discrimination, either because they are ‘trained’ using data that already contained such biases, or because other information and data (from geography, social media, spending patterns, etc.) generate results that are highly correlated with race or gender.

The **business models of some incumbent financial institutions may be undermined** by competition from new entrants; an inability to keep pace with technology; being held back by legacy systems that are expensive to maintain, expensive to replace entirely, and difficult to run alongside more modern systems; and an inability to generate cost savings and economies of scale.

Some **new entrants have failed and may fail**, including as a result of trying to grow too rapidly; automating a flawed process; taking on risks that other financial institutions may have good reasons not to take on; and not having an approach and mindset centred on the importance of good governance, risk management, financial soundness, and long-term stability.

## Risks to financial stability

The use of AI and ML could pose risks to financial stability<sup>22</sup>, although this is unlikely to be a significant concern in most emerging market and developing economies, given the current low usage of AI and ML.

---

<sup>21</sup> See Toronto Centre (2021a) for a discussion of operational resilience.

<sup>22</sup> These risks are discussed in more detail in Financial Stability Board (2017).

**Herd-like behaviours**, for example where lenders or insurers use similar approaches to credit and insurance underwriting risk assessments, or where trading firms use and follow similar algorithms, could generate closely correlated risks and sharp movements in asset prices. If AI and ML are perceived to outperform other underwriting and trading strategies this could result in more financial institutions adopting similar strategies. And if a critical segment of financial institutions relies on the same data sources and algorithmic strategies, then under certain market conditions a shock to those data sources could affect that segment as if it were a single institution. Equally, however, the use of unstructured data in AI and ML models may lead to the development of different, less correlated, models which would reduce the risk of herd-like behaviours.

**New and unexpected forms of interconnectedness between financial markets and institutions** could arise from the use of AI and ML, for example through the use by various institutions of previously unrelated data sources. Greater interconnectedness could amplify shocks.

**Trading algorithms** can generate sharp movements in securities prices<sup>23</sup>; could potentially be vulnerable to adaptive behaviours across ML models<sup>24</sup>; and could potentially be vulnerable to attacks by insiders or cyber-criminals seeking to manipulate market prices by feeding false information into automated trading strategies.<sup>25</sup>

**Economies of scale in the use of AI, ML and big data** could lead to market concentrations and monopolies, from incumbent players becoming even larger, from the entry and dominance of Bigtech companies, or from the rapid growth of new entrants. In addition, the increasing concentration on a small number of large and globally active third-party providers (for example, in the provision of cloud computing<sup>26</sup>) could lead to the emergence of new systemically important players that could fall outside the regulatory perimeter.

**Alternative channels of financial intermediation** could emerge from the success of non-traditional players, such as non-bank lenders. These channels may fall outside the regulatory perimeter or be weakly regulated and supervised.

## Risks to consumers

In addition to the risks to consumers arising from the failure of financial institutions and financial instability, consumers may face various other risks from the use of AI and ML by financial institutions. Some of these risks may be even higher in those emerging economies where supervisory and regulatory capacity is low and where legal frameworks are weak.

**Mis-selling** – AI and ML offer new ways for financial institutions to mis-treat their customers, for example pressure selling of poor or unsuitable products through digital channels and social media platforms, and using actual or apparent complexity to increase the asymmetry of information and understanding between financial institutions and their customers.

**Mis-advising** – innovations such as robo-advice are only as good as the data and programming on which they operate. Financial institutions can bias the process in their

---

<sup>23</sup> For example, the sharp intraday decline of the Dow Jones index on 6 May 2010 was attributed to a large sell order generated by an automated execution algorithm. See <https://www.sec.gov/files/marketevents-report.pdf>

<sup>24</sup> Self-learning and deep learning models could recognize mutual interdependencies and adapt to the behaviour and actions of other market participants or other machine learning models, which could potentially deliver the same outcome as collusive behaviour, without any human intervention and perhaps without the owners of the models even being aware of it.

<sup>25</sup> For example, there were market moves across equities, bonds, foreign exchange, and commodities in April 2013 after trading algorithms reacted to a fraudulent news Tweet announcing two explosions at the White House. See Financial Stability Board (2017).

<sup>26</sup> See Toronto Centre (2020) for a discussion of cloud computing.

favour by programming computers to recommend their own products and services, even when this is not the most suitable outcome for consumers. A “better customer experience” is by no means a guaranteed outcome from the use of AI and ML.

**Embedded bias** – biased training data taken from existing biased processes and data sets (for example, incomplete or unrepresentative data, or data that reflect prevailing prejudices) will teach AI and ML models to be biased as well, and therefore unfair to consumers.

**Financial exclusion** – even if, overall, innovations in the use of technology and data should increase access to financial products and services, some consumers could find that credit and insurance cover becomes more difficult to obtain or more expensive. The use of new alternative data sources, such as online behaviour or non-traditional financial and other information, could generate model predictions that make some consumers less attractive to banks and insurers than had been the case before the use of AI and ML models.

**Lack of “explainability” of ML models** – consumers will find it even more difficult to understand the basis of decisions made about them when these decisions are based on AI and ML.

**Data privacy and security** – consumers are vulnerable to the loss and misuse of data, and are unlikely to understand fully the ways in which their financial and non-financial data are being used.

**Fraud and scams** – the use of digital channels increases the scope for attempted frauds and scams on unsuspecting and vulnerable consumers.

**Cyber attacks and other operational resilience failures** – operational resilience failures at financial institutions may limit the ability of customers to access services, products and even their cash and other assets; and expose them to various types of fraud and to the loss or manipulation of data.

**Complaints handling** – new entrants may not have the organizational structure to handle complaints effectively; digital communications make it more difficult to speak to a real person; and the increasing fragmentation of product and service providers (for example, through open banking and a more modular approach to banking and insurance) can lead to a lack of clarity as to who is responsible when things go wrong.

**Increasing concentration among a small number of large financial institutions** – consumers can be harmed by higher prices and a reduced choice of products and services.

## Regulation and supervision: way forward

Supervisory authorities need to ensure that the use of AI and ML by financial institutions is consistent with their supervisory objectives (the safety and soundness of supervised financial institutions, consumer protection, financial stability, market integrity and fair, orderly and transparent markets, anti-money laundering, financial inclusion, etc), while also supporting innovation and competition in the financial sector.

A supervisory authority may face some difficult trade-offs here. For example, as discussed above, the use of AI and ML in credit scoring may enhance financial inclusion while at the same time introducing new risks for financial institutions, financial stability and consumer protection.

Emerging risks from the deployment of AI and ML therefore need to be identified and (where necessary) mitigated to support the responsible use of AI and ML. Within a proportional and risk-based approach, supervisory authorities need to consider:

- Where are the biggest risks from the use of AI and ML in the financial sector, and how significant are they?
- What can be done to control and mitigate these risks?
- Are new regulations (international standards, national laws, regulatory rules, supervisory guidance) required, or can existing approaches be adapted to cover these risks?
- What are the implications for supervisory resources – numbers, skills and expertise?

In addressing these questions, a supervisory authority might usefully focus on three complementary regulatory and supervisory responses to the increasing use of AI and ML in the financial sector– the formulation of high-level principles for ‘trustworthy’ AI; the extension of requirements for the use of models by financial institutions; and some specific examples of regulation and supervision in this area.

### High-level principles for ‘trustworthy’ AI

Several sets of high-level principles and guidelines have emerged to govern the use of AI and ML generally, not just in the financial sector, with the objective of delivering ‘trustworthy AI’. Many of these principles and guidelines are directly relevant to supervisory authorities in the financial sector, and could therefore form the basis of an amended and expanded set of requirements for the use of AI and ML by financial institutions.

#### **OECD principles for trustworthy AI**

The OECD (2019) recommends that governments should develop policies, in co-operation with all stakeholders, to promote trustworthy AI systems and achieve fair and beneficial outcomes, based on five broad principles:

**Investing in responsible AI research and development**, not just to encourage innovation and competition but also to address the ethical, legal, and social implications of AI, including bias, privacy, transparency, accountability and the safety of AI, and difficult technical challenges such as explainability.

**Fostering an enabling digital ecosystem for AI**, including government investment in, and providing incentives to the private sector to invest in, AI enabling infrastructure and technologies such as high-speed broadband, computing power and data storage; and encouraging the sharing of AI knowledge through mechanisms such as open AI platforms and data sharing frameworks while respecting privacy, intellectual property and other rights.

**Providing an agile and controlled policy environment for AI**, by reviewing existing laws, regulations, policy frameworks and assessment mechanisms as they apply to AI and adapt them, or develop new ones as appropriate; and by establishing mechanisms for the continuous monitoring, reporting, assessing and addressing of the implications of AI systems that may pose significant risks or target vulnerable groups.

**Building human capacity and preparing for job transformation**, which would include the building of new skills and technological literacy in the financial sector.

**International cooperation for trustworthy AI**, recognizing that trustworthy AI is a global challenge and the value of international and cross-sectoral collaboration, knowledge-sharing, and standards.

## European Union guidelines for trustworthy AI

A European high level expert group (2019) established by the European Commission identified three key components of trustworthy AI, namely that it should be:

- **lawful**, complying with all applicable laws and regulations;
- **ethical**, ensuring adherence to ethical principles and values; and
- **robust**, from both a technical and a social perspective, since, even with good intentions, AI systems can cause unintentional harm.

To achieve this, the working group proposed guidelines on:

- ✓ Developing, deploying, and using AI systems in a way that adheres to the ethical principles of respect for human autonomy, prevention of harm, fairness and explainability.
- ✓ Ensuring that the development, deployment, and use of AI systems meets the requirements for trustworthy AI - human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being, and accountability.
- ✓ Paying particular attention to situations involving more vulnerable groups that have historically been disadvantaged or are at risk of exclusion.
- ✓ Acknowledging that, while bringing substantial benefits to individuals and society, AI systems also pose risks and may have a negative impact.
- ✓ Adopting adequate measures to mitigate these risks when appropriate, and proportionately to the magnitude of the risk.
- ✓ Fostering research and innovation to help assess AI systems and to further the achievement of trustworthy AI.
- ✓ Communicating, in a clear and proactive manner, information to stakeholders about an AI system's capabilities and limitations, enabling realistic expectation setting.
- ✓ Facilitating the traceability and auditability of AI systems, particularly in critical contexts or situations.
- ✓ Involving stakeholders throughout an AI system's life cycle.
- ✓ Adopting a trustworthy AI assessment list when developing, deploying or using AI systems.
- ✓ Ensuring that trustworthy AI is not about ticking boxes, but about continuously identifying and implementing requirements, evaluating solutions, ensuring improved outcomes throughout an AI system's lifecycle, and involving stakeholders in this.

In part in response to these recommendations, the European Commission (2021) has proposed legislation covering the use of AI across all sectors of the economy. This proposes a risk-based approach, which would ban some uses of AI, and apply tougher regulation for riskier uses.<sup>27</sup>

Providers of higher-risk AI systems would be required to:

- Comply with rules on data and data governance; documentation and record-keeping; transparency and provision of information to users; human oversight; and robustness, accuracy and security.
- Undertake ex-ante conformity assessments to establish that high-risk AI systems meet these requirements before they can be put into service.
- Undertake 'in use' monitoring to detect problems and to mitigate them.

---

<sup>27</sup> In the financial sector, credit scoring is specified in the EU proposals as a high-risk use of AI.

These principles and guidelines provide a useful starting point, but supervisory authorities will need to consider how each principle could be applied in practice. The principles need to be converted into some form of regulatory requirements and/or supervisory expectations that describe what counts as meeting each principle.

Some of these principles and guidelines are already applied by financial regulation and supervision, for example in the use of models by financial institutions, and by the “fair treatment of consumers” requirements imposed by some supervisory authorities with responsibility for retail conduct. In these cases, existing requirements may need to be clarified and adjusted to reflect the ways in which AI and ML applications may affect risks that are already familiar to financial regulators and supervisors.

But some of the principles extend beyond the traditional scope of regulation and supervision. These may require the development of new, or substantially amended, regulatory requirements and supervisory expectations. Three extensions are of particular importance to supervisors here.

First, the ‘black-box’ nature of ML models makes it more difficult for financial institutions to understand how a model works and to explain this to their customers and to their supervisors. Bank of England (2109) discusses how some progress on ‘explainability’ can be made by studying the inputs and the outcomes of a ML model, but not its inner workings. Various statistical analyses can be used to assess which drivers (explanatory variables) are determining the outcomes, and therefore to begin to explain the results of the model.

Second, ML models are also more difficult to validate, and in the absence of clear causal relationships (between the explanatory variables and the dependent variable) it may be more difficult to identify when a model is beginning to perform less well and to respond to this by respecifying the model. ML models may break down quickly and seriously when previous patterns change.

Third, as discussed above, AI and ML can reinforce biases and discrimination and potentially lead to undesirable outcomes. This raises issues relating to fairness and to ‘machine ethics’, namely the imposition of ethical norms on the behaviour of artificial agents. There are also ethical questions regarding the use of sensitive personal data, for example using medical information (or proxies for this, such as social media and dietary habits revealed by spending data) in insurance underwriting decisions.

The practical application of these principles therefore requires a supervisory authority to determine what would count as acceptable levels of explainability, robustness, and transparency, and what counts as a sufficiently fair and unbiased use of AI and ML, and to impose requirements accordingly on financial institutions using AI and ML applications.

One example of a translation of high-level principles for trustworthy AI into more specific regulatory and supervisory requirements are the Monetary Authority of Singapore principles to promote fairness, ethics, accountability, and transparency in the use of AI and data analytics.

**Monetary Authority of Singapore (MAS) principles to promote fairness, ethics, accountability, and transparency in the use of AI and data analytics**

The MAS (2018) set out and amplified four high-level principles for the use of AI and data analytics (AIDA):

**Fairness**

- Individuals or groups of individuals are not systematically disadvantaged through AIDA-driven decisions unless these decisions can be justified.

- The use of personal attributes as input factors for AIDA-driven decisions should be justified.
- Data and models used for AIDA-driven decisions are regularly reviewed and validated for accuracy and relevance, and to minimize unintentional bias.
- AIDA-driven decisions are regularly reviewed so that models behave as designed and intended.

### **Ethics**

- The use of AIDA is aligned with the firm's ethical standards, values and codes of conduct.
- AIDA-driven decisions are held to at least the same ethical standards as human-driven decisions.

### **Accountability**

#### *Internal Accountability*

- Use of AIDA in AIDA-driven decision-making is approved by an appropriate internal authority.
- Firms using AIDA are accountable for both internally developed and externally sourced AIDA models.
- Firms using AIDA should proactively raise senior management and board awareness of their use of AIDA.

#### *External Accountability*

- Data subjects are provided with channels to enquire about, submit appeals for, and request reviews of AIDA-driven decisions that affect them.
- Verified and relevant supplementary data provided by data subjects are taken into account when performing a review of AIDA-driven decisions.

### **Transparency**

- To increase public confidence, use of AIDA is proactively disclosed to data subjects as part of general communication.
- Data subjects are provided, upon request, with clear explanations on what data are used to make AIDA-driven decisions about the data subject, and how the data affect the decision.
- Data subjects are provided, upon request, with clear explanations on the consequences that AIDA-driven decisions may have on them.

## **Model requirements**

Another building block for developing regulatory and supervisory responses to the increasing use of AI and ML in the financial sector is to consider how the existing requirements on the use of traditional statistical models by financial institutions could be adjusted to incorporate AI and ML.

The main challenges here relate to the complexity of ML models; the difficulties in interpreting their results; and the difficulties for both management and supervisors in understanding these models and their outcomes.

The Financial Stability Institute (2021) has set out a helpful taxonomy of AI principles and their application to AI and ML based models, comparing these with supervisory approaches to traditional models used by financial institutions.<sup>28</sup>

<sup>28</sup> The table below is based in part on Financial Stability Institute (2021). See also Toronto Centre (2021b).

AI Principle	Supervisory considerations for AI and ML applications
Reliability and soundness	<p>Similar supervisory expectations as those for traditional models - model validation, defining metrics of accuracy, updating/retraining of models, data quality, internal and supervisory review.</p> <p>In addition, there may be a need to assess the reliability and soundness of AI and ML applications relevant to retail and wholesale conduct, including an assessment of adherence to relevant protocols on data privacy, conduct requirements, and cybersecurity.</p>
Accountability and governance	<p>Similar supervisory expectations as outlined in general accountability, governance and oversight requirements.</p> <p>In addition, for AI and ML applications, there should be a greater supervisory emphasis on:</p> <ul style="list-style-type: none"> <li>• the necessity of human involvement and oversight;</li> <li>• the importance of board and senior management understanding and control of AI and ML applications;</li> <li>• placing accountability firmly and clearly on responsible persons within a financial institution; and</li> <li>• ‘external accountability’ in the sense of prospective or existing customers being aware of AI and ML based decisions, and having channels for recourse.</li> </ul>
Transparency	<p>Similar supervisory expectations as for traditional models.</p> <p>In addition, for AI and ML applications there should be a greater supervisory emphasis on ‘explainability’ and auditability – internal and supervisory reviews need to be able to ‘unpack’ the black box.</p> <p>Also, for customer-facing AI and ML applications there should be external disclosure to data subjects on the data used to drive decisions and on how the data affect those decisions.</p>
Fairness	<p>Fairness expectations are not typically applied to traditional models used for prudential purposes.</p> <p>For AI and ML applications, supervisory expectations on fairness should relate to addressing or preventing biases in AI and ML applications that could lead to discriminatory outcomes. There is a need for adequate testing and ‘training’ of applications with unbiased data and feedback mechanisms.</p>
Ethics	<p>Not typically applied to traditional models.</p> <p>Ethics considerations are broader than “fairness” and relate to ensuring that customers will not be exploited or harmed, accidentally or deliberately, through bias, discrimination or other causes.</p> <p>Need to test and sense check the outcomes of AI and ML applications for fairness and ethical use.</p>

Similarly, the European Banking Authority issued a discussion paper (2021) on additional requirements that supervisors may require banks to meet if they are to be allowed to use ML models to calculate capital requirements.

#### **European Banking Authority (2021) principle-based recommendations on using ML models for prudential purposes**

- Appropriate levels of knowledge of the model's functioning for different levels of staff in the bank.
- Avoiding unnecessary complexity in the modelling approach if it is not justified by a significant improvement in the predictive capacities.
- Statistical analysis of the relationship of each single risk driver (explanatory variable) with the output variable, and the overall weight of each risk driver in determining the output variable, to detect which risk drivers influence model prediction the most.
- Assessing the economic relationship of each risk driver with the output variable to ensure that the model estimates are plausible and intuitive.
- A summary document in which the model is clearly explained, based on the analyses described above.
- Ensure that potential biases in the model (for example, overfitting to the training sample) are detected.
- The importance of human judgement.
- Where ML models are frequently updated, analysis of the reason for such regular updates.
- For complex ML models with limited explainability or for frequently updated models, a reliable validation is particularly important and might require increased depth and/or frequency.

### Examples

A third building block for developing regulatory and supervisory responses to the increasing use of AI and ML in the financial sector is to consider international and national standards for AI and ML applications.

This section summarizes standards from IOSCO, the UK Prudential Regulation Authority, and the European Banking Authority. Most of these standards are generally applicable across sectors and across various AI and ML applications.

**IOSCO** (2021) recommends six “measures” requiring financial institutions using AI and ML to have in place:

**Governance** – responsibility for the oversight of the development, testing, deployment, monitoring and controls, and updating of AI and ML applications should be assigned to designated senior managers with the relevant skill sets and knowledge.

Internal model governance committees or model review boards of financial institutions should be tasked with the setting of governance standards and processes for model building, documentation, and validation for any AI and ML application.

**Testing and monitoring** – financial institutions should test and monitor the algorithms to validate the results of each AI and ML application on a continuous basis, to ensure the application (a) behaves as expected in stressed and unstressed market conditions, and

(b) operates in a way that complies with regulatory obligations (including on fairness and ethical outcomes).

**Adequate skills, expertise and experience** – to develop, test, deploy, monitor and oversee the controls over AI and ML applications, compliance and risk management functions should be able to understand and challenge the algorithms that are produced and conduct due diligence on any third-party provider.

**Outsourcing controls** – financial institutions should understand their reliance and manage their relationship with third-party providers, including monitoring their performance and conducting oversight.

Clear service level agreements and contracts should be in place clarifying the scope of the outsourced functions and the responsibility of the service provider, with clear performance indicators and rights and remedies for poor performance.

**Disclosure** – financial institutions should disclose meaningful information to customers and clients around their use of AI and ML where this has an impact on client outcomes.

Supervisors should consider what information they may require from financial institutions using AI and ML so that they can exercise supervisory oversight.

**Data quality** – appropriate controls should be in place to ensure that data are of sufficient quality to prevent biases and sufficiently broad for a well-founded application of AI and ML.

The **UK Prudential Regulation Authority** (2018) established a set of supervisory expectations for algorithmic trading that are very similar to the six IOSCO measures, demonstrating the general applicability of the IOSCO measures:

**Governance** – financial institutions should have a clearly documented trading policy, clear lines of responsibility, good understanding and ‘explainability’ of their application of AI and ML to algorithmic trading, and proper oversight.

**Approval process** – covering the ownership, testing, risk assessment, and sign-off of each AI and ML application.

**Testing and deployment processes** – covering the design, validation, implementation, and subsequent variations of each AI and ML application.

**Inventories and documentation** – proper documentation of the algorithms, controls, and trading systems architecture.

**Risk management and other controls** – the understanding of each application by the business owners and by risk management and compliance; the adequacy of risk identification; a full range of controls; and the operation of a “kill switch” (to quickly shut down an AI-based system if it ceases to function according to its intended purpose).

The **European Banking Authority** (2022) has published proposals on the regulatory and supervisory response to non-bank lenders, including lenders using AI and ML for credit scoring. This usefully supplements the model requirement considerations discussed above by focusing on:

**Regulatory perimeter** – supervisory authorities need to consider whether non-bank lenders should be included within the scope of regulation, and if so what legislative changes may be required to achieve this.

**Micro-prudential risks** – lending activities carry credit, liquidity and operational risks, the nature and magnitude of which may change through the use of AI and ML.

**Macro-prudential risks** – although currently relatively small in most countries, non-bank lending could generate macro-prudential risks from over-borrowing, contagion, and regulatory arbitrage.

**Consumer protection** – in addition to the more general consumer protections against inadequate product disclosures (especially in a digital context), mis-selling (for example, through aggressive or misleading digital marketing), inadequate complaint handling procedures, and insufficient safeguards against over-indebtedness, the use of AI and ML also requires consumer protections relating to:

- consumer understanding of what data are being used for credit scoring and pricing purposes;
- ‘explainability’ – consumers should have the right to obtain human intervention with the aim of correcting or explaining the outcome of the credit assessment made, including the main variables used in the process; and
- avoiding embedded biases and discrimination.

## Suptech

Financial institutions can use AI and ML to deliver compliance with regulatory requirements (a subset of “Regtech”) – for example through applications relating to risk management, anti-money laundering and fraud controls, regulatory reporting, and scanning texts for compliance issues.<sup>29</sup>

There is also scope for supervisory authorities to use AI and ML to improve the effectiveness and efficiency of supervision (“Suptech”).<sup>30</sup>

Some supervisors have made use of technological innovations to collect and analyze information. This enables supervisors to go beyond the manual receipt and filing of regulatory reports, supplemented by some limited spreadsheet analysis, to the digitization and electronic transfer of at least some elements of the reporting and analysis process, and to a more sophisticated analysis of these data by the supervisory authority.

AI and ML can support the real-time capture and analysis of large volumes of data from financial institutions, enabling supervisors to supplement standard calculations with the identification of outliers and anomalies, the identification of patterns and trends (within and across financial institutions), the creation of early warning indicators, network analysis, and checks on the quality of the data reported by financial institutions.

---

<sup>29</sup> See Toronto Centre (2017).

<sup>30</sup> For more details on Suptech, see the series of papers from Toronto Centre (2017, 2018, and 2021d), the Financial Stability Institute (2018, 2019a, 2019b, 2020, and 2021b), and the Financial Stability Board (2020).

AI and ML can also be used by supervisors to detect collusive behaviour, price manipulation and fraud in securities and other traded markets, using data supplied by financial institutions and financial market infrastructure (exchanges, and clearing and settlement systems).

AI and ML can also be used to analyze customers and customer transactions (these data could be supplied to the supervisory authority by financial institutions) to identify possible cases of money laundering, and possible cases where financial institutions may have classified customers incorrectly (for example, the analysis undertaken by a supervisory authority may suggest that a customer is high risk, when a financial institution has classified the customer as low risk under its own know-your-customer procedures).

There is also scope to extend supervisory analysis beyond the data contained in traditional regulatory reports from financial institutions. For example, supervisors could use AI and ML to analyse:

- Published documents from financial institutions – annual reports, ESG reports, product literature, advertising, customer contracts.
- Unpublished documents from financial institutions – board (and board sub-committee) minutes, ICAAP and ORSA reports, credit and underwriting files.
- Social media references to financial institutions – complaints, sentiment analysis.
- Inter-connectedness among board members, and other ‘fit and proper’ checks.

As with the use of AI and ML by financial institutions, a supervisory authority needs to consider:

- Whether it has the necessary **skills, expertise, knowledge and resources** to implement any Suptech application. In some emerging market and developing economies this may be more likely where the central bank is the supervisor – a central bank may have more advanced capabilities in data and data analysis, and is more likely to have the resources to invest in Suptech applications.
- Whether supervisors **understand what AI and ML can - and cannot - deliver**. Suptech may provide opportunities, but it also has limitations.
- **How supervisors and technical experts can work effectively together** to identify and develop Suptech opportunities that will genuinely enhance and add value to supervision, or at least make it more efficient. Suptech needs to be directed towards answering the right questions, not the application of technology for its own sake.
- Whether the data and other information are available and are of sufficiently **good quality**.
- What supervisors will **do with the results of Suptech**. Suptech may provide useful inputs to supervision, but it still requires human judgement in interpreting the results and in using them to drive risk-based, proportional, efficient and effective supervision. For example, Suptech may help to identify higher risk financial institutions, or higher risk activities within financial institutions, or system-wide issues, all of which may provide a basis for further off-site analysis and for the on-site questioning of senior management, board members, heads of business units, and heads of risk and control functions.<sup>31</sup>

---

<sup>31</sup> See Toronto Centre (2022) for a discussion of the importance of using on-site supervision to ask open-ended questions to board members and senior management to assess the effectiveness of a financial institution’s corporate governance.

## Conclusion

Financial institutions are increasing their use of AI and ML across a wide range of applications, including to improve customer service, for credit and insurance underwriting, to identify suspicious transactions, for the provision of advice to consumers, and for trade finance. These applications bring both opportunities and risks.

This Note has set out the risks to financial institutions, financial stability and consumers from the use of AI and ML. Some of these risks can be viewed as the impact of the use of AI and ML on traditional risks. Other risks are new, in particular those relating to the explainability and robustness of models, fairness, and ethics. Supervisory authorities need to respond when these risks become significant, given their prudential, financial stability, retail and wholesale market conduct, anti-money laundering and financial inclusion objectives.

This Note has also provided some frameworks for considering how a supervisory authority might respond to these risks, taking account of the need for a supervisory authority to:

- identify and assess the significance of risks from the use of AI and ML in the financial sector;
- consider what financial institutions might be required to do to control and mitigate these risks;
- consider whether new regulations and supervisory guidance are required, and how existing approaches might need to be adapted to cover these risks; and
- assess the implications for supervisory resources – numbers, skills and expertise.

Supervisory authorities should also consider whether they can themselves make good use of AI and ML to improve the effectiveness and efficiency of supervision.

## References

- Bank of England. [Machine learning explainability in finance: an application to default risk analysis](#). Staff Working Paper 816. August 2019.
- European Banking Authority. [Discussion paper on machine learning for IRB models](#). November 2021.
- European Banking Authority. [Final Report on response to the non-bank lending request from the CfA on digital finance](#). April 2022.
- European Commission. [Regulatory framework proposal on artificial intelligence](#). April 2021.
- European Union High-Level Expert Group on Artificial Intelligence. [Ethics guidelines for trustworthy AI](#). April 2019.
- Financial Stability Board. [Artificial intelligence and machine learning in financial services](#). November 2017.
- Financial Stability Board. [The use of supervisory and regulatory technology by authorities and regulated institutions](#). October 2020.
- Financial Stability Institute. [Innovative technology in financial supervision \(suptech\) – the experience of early users](#). July 2018.
- Financial Stability Institute. [Suptech applications for anti-money laundering](#). August 2019a.
- Financial Stability Institute. [The suptech generations](#). October 2019b.
- Financial Stability Institute. [From data reporting to data-sharing: how far can suptech and other innovations challenge the status quo of regulatory reporting?](#) December 2020.
- Financial Stability Institute. [Humans keeping AI in check – emerging regulatory expectations in the financial sector](#). August 2021a.
- Financial Stability Institute. [Suptech tools for prudential supervision and their use during the pandemic](#). December 2021b.
- International Monetary Fund. [FinTech in financial inclusion: Machine learning applications in assessing credit risk](#). IMF Working Paper WP/19/109. May 2019.
- International Monetary Fund. [Powering the digital economy: Opportunities and risks of artificial intelligence in finance](#). September 2021.
- International Organisation of Securities Commissions. [The use of artificial intelligence and machine learning by market intermediaries and asset managers](#). September 2021.
- Monetary Authority of Singapore. [Principles to Promote Fairness, Ethics, Accountability and Transparency \(FEAT\) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector](#). November 2018.
- O'Keefe, John P. [The Use of Machine Learning for Bank Financial Stress Tests](#). January 2022.
- Organisation for Economic Co-operation and Development. [Scoping the OECD AI principles: Deliberations of the expert group on artificial intelligence at the OECD](#). November 2019.

Organisation for Economic Co-operation and Development. [The impact of big data and artificial intelligence in the insurance sector](#). January 2020.

Organisation for Economic Co-operation and Development. [Artificial intelligence, machine learning and big data in finance: Opportunities, challenges, and implications for policy makers](#). August 2021.

Prudential Regulation Authority. [Algorithmic trading](#). June 2018.

Toronto Centre. [FinTech, RegTech and SupTech: What They Mean for Financial Supervision](#). August 2017.

Toronto Centre. [SupTech: Leveraging technology for better supervision](#). July 2018.

Toronto Centre. [Cloud Computing: Issues for Supervisors](#). November 2020.

Toronto Centre. [Operational Resilience: The Next Frontier for Supervisors?](#) April 2021a.

Toronto Centre. [Supervision of Bank Model Risk Management](#). July 2021b.

Toronto Centre. [A Climate Risk Toolkit for Financial Supervisors](#). September 2021c.

Toronto Centre. [Using SupTech to Improve Supervision](#). November 2021d.

Toronto Centre. [Supervising Corporate Governance: Pushing the Boundaries](#). January 2022.

World Bank International Finance Corporation. [Artificial intelligence in emerging markets: Opportunities, trends, and emerging business models](#). March 2021.