1. Introduction

This paper explores the use of artificial intelligence (AI) and machine learning (ML) in the financial sector and the resultant policy implications.[1] It provides a nontechnical background on the evolution and capabilities of AI/ML systems, their deployment and use cases in the financial sector, and the new challenges they present to financial sector policymakers.

AI/ML systems have made major advances over the past decade. Although the development of a machine with the capacity to understand or learn any intellectual task that a human being performs is not within immediate grasp, today's AI systems can perform quite well tasks that are well defined and normally require human intelligence. The learning process, a critical component of most AI systems, takes the form of ML, which relies on mathematics, statistics, and decision theory. Advances in ML and especially in deep learning algorithms are responsible for most of the recent achievements, such as self-driving cars, digital assistants, and facial recognition.[2]

The financial sector, led by financial technology (fintech) companies, has been rapidly increasing the use of AI/ML systems (Box 1). Recent adoption by the financial sector of technological advances, such as big data and cloud computing, coupled with the expansion of the digital economy, made the effective deployment of AI/ML systems possible. A recent survey of financial institutions (WEF 2020) shows that 77 percent of all respondents anticipate that AI will be of high or very high overall importance to their businesses within two years. McKinsey (2020a) estimates the potential value of AI in the banking sector to reach $1 trillion.

AI/ML capabilities are transforming the financial sector.[3] AI/ML systems are reshaping client experiences, including communication with financial service providers (for example, chat bots), investing (for example, robo-advisor), borrowing (for example, automated mortgage underwriting), and identity verification (for example, image recognition). They are also transforming the operations of financial institutions, providing significant cost savings by automating processes, using predictive analytics for better product offerings, and providing more effective risk and fraud management processes and regulatory compliance. Finally, AI/ML systems provide central banks and prudential oversight authorities with new tools to improve systemic risk surveillance and strengthen prudential oversight.

The COVID-19 pandemic has further increased the appetite for AI/ML adoption in the financial sector. BoE (2020) and McKinsey (2020b) find that a considerable number of financial institutions expect AI/ML to play a bigger role after the pandemic. Key growth areas include customer relationship and risk management. Banks are exploring ways to leverage their experience of using AI/ML to handle the high volume of loan applications during the pandemic to improve their underwriting process and fraud detection. Similarly, supervisors relying on off-site intensive supervision activities during the pandemic could further explore AI/ML-supported tools and processes in the post-pandemic era.

The rapid progress in AI/ML development could deepen the digital divide between advanced and developing economies. AI/ML deployment, and the resulting benefits, have been concentrated largely in advanced economies and a few emerging markets. These technologies could also bring significant benefits to developing economies, including enhanced access to credit by reducing the cost of credit risk assessments, particularly in countries that do not have an established credit registry (Sy and others 2019). However, these economies are falling behind, lacking

[1] Following the Oxford Dictionary, AI is defined as the theory and development of systems able to perform intellectual tasks that usually require human intelligence. ML is the learning component of an AI system, and is defined as the process that uses experience, algorithms, and some performance criterion to get better at performing a specified task. Given that AI and ML heavily overlap and that most statements in this paper hold true for both concepts, the terms are often used as a pair (AI/ML).

[2] See Annex 1 for more details.

[3] This includes revenue gains and cost savings.

the necessary investment, access to research, and human capital.[4] Bridging this gap will require developing a digital-friendly policy framework anchored around four broad policy pillars: investing in infrastructure; investing in policies for a supportive business environment; investing in skills; and investing in risk management frameworks (IMF 2020).

Cooperation among countries and between the private and public sectors could help mitigate the risk of a widening digital divide. So far, global initiatives—including the development of principles to mitigate ethical risks associated with AI (UNESCO 2021; OECD 2019), calls for cooperation on investing in digital infrastructure (see, for example, Google and International Finance Corporation (2020)), and the provision of access to research in low-income countries (see, for example, AI4Good.org)—have been limited. Multilateral organizations could play an important role in transferring knowledge, raising investments, building capacity, and facilitating a peer-learning approach to guide digital policy efforts in developing economies. Similarly, the membership in several intergovernmental working groups on AI (such as the Global Partnership on Artificial Intelligence and the OECD Network of Experts on AI, among others) could be expanded to include less-developed economies.

AI/ML adoption in the financial sector is bringing new unique risks and challenges that need to be addressed to ensure financial stability. AI/ML-based decisions made by financial institutions may not be easily explainable and could potentially be biased. AI/ML adoption brings in new unique cyber risks and privacy concerns. Financial stability issues could also arise with respect to the robustness of the AI/ML algorithms in the face of structural shifts and increased interconnectedness through widespread reliance on few AI/ML service providers. Chapter 2 explores the adoption of AI/ML in the financial sector and possible associated risks, Chapter 3 discusses related policy concerns, and Chapter 4 provides some conclusions.

Box 1. Artificial Intelligence and Machine Learning Capabilities

- *Forecasting.* Machine learning algorithms are used for forecasting and benefit from using large data sets. They usually perform better than traditional statistical or econometric models.[1] In the financial sector, this is used in such areas as credit risk scoring, economic and financial variables forecasting, risk management, and so on.
- *Natural language processing.* Artificial intelligence systems can communicate by understanding and generating human language. Boosted by deep learning and statistical models, natural language processing has been used in the financial sector in such applications as chat bots, contract reviewing, and report generation.
- *Image recognition.* Facial and signature recognition is being used by some financial institutions and financial technology companies to assist with carrying out certain anti-money laundering/combating the financing of terrorism (AML/CFT) requirements (for example, the identification and verification of customers for customer due diligence process), and for strengthening systems security.
- *Anomaly detection.* Classification algorithms can be applied to detect rare items, outliers, or anomalous data. In the financial sector, insider trading, credit card and insurance fraud detection, and AML/CFT are some of the applications that leverage this capability (Chandola, Banerjee, and Kumar 2009).

[4]See Alonso and others (2020) for a broader discussion about possible implications of AI on developing economies. In particular, the paper finds that the new technology risks widening the gap between rich and poor countries by shifting more investment to advanced economies where automation is already established, with negative consequences for jobs in developing economies.

2. Artificial Intelligence in the Financial Sector

The capability of acquiring large sets of data from the environment and processing it with artificial intelligence (AI) and machine learning (ML) is changing the financial sector landscape. AI/ML facilitates enhanced capacity to predict economic, financial, and risk events; reshape financial markets; improve risk management and compliance; strengthen prudential oversight; and equip central banks with new tools to pursue their monetary and macroprudential mandates.

A. Forecasting

AI/ML systems are used in the financial sector to forecast macro-economic and financial variables, meet customer demands, provide payment capacity, and monitor business conditions. AI/ML models offer flexibility compared to traditional statistical and econometric models, can help explore otherwise hard-to-detect relationships between variables, and amplify the toolkits used by institutions. Evidence suggests that ML methods often outperform linear regression-based methods in forecast accuracy and robustness (Bolhuis and Rayner 2020).

While the use of AI/ML in forecasting offers benefits, it also poses challenges. Use of nontraditional data (for example, social media data, browsing history, and location data) in AI/ML could be beneficial in finding new relationships between variables. Similarly, by using AI natural language processing (NLP), unstructured data (for example, the information in email texts) can be brought into the forecasting process. However, the use of nontraditional data in financial forecasting raises several concerns, including the governing legal and regulatory framework; ethical and privacy implications; and data quality in terms of cleanliness, accuracy, relevancy, and potential biases.

B. Investment and Banking Services

In the financial sector, advances in AI/ML in recent years have had their greatest impact on the investment management industry. The industry has used technology for decades in trading, client services, and back-office operations, mostly to manage large streams of trading data and information and to execute high-frequency trading. However, AI/ML and related technologies are reshaping the industry by introducing new market participants (for example, product customization), improved client interfaces (for example, chatbots), better analytics and decision-making methods, and cost-reduction through automated processes (Box 2).

Compared to the investment management industry, the penetration of AI/ML in banking has been slower. The banking industry has traditionally been at the forefront of technological advancements (for example, through the introduction of ATMs, electronic card payments, and online banking). However, confidentiality and the proprietary nature of banking data have slowed AI/ML adoption. Nonetheless, AI/ML penetration in the banking industry has accelerated in recent years, in part on account of rising competition from financial technology (fintech) companies (including fintech lenders), but also fueled by AI/ML's capacity to improve client relations (for example, through chatbots and AI/ML-powered mobile banking), product placement (for example, through behavioral and personalized insights analytics), back-office support, risk management, credit underwriting (Box 3), and, importantly, cost savings.[5]

[5]The aggregate potential cost savings for banks from AI/ML systems is estimated at $447 billion by 2023 (Digalaki 2021).

Box 2. Artificial Intelligence in Investment Management—Sample Use Cases[1]

- Increased market liquidity provision through a wider use of high-frequency algorithmic trading and more efficient market price formation.
- Expanded wealth advisory services by providing personal and targeted investment advice to mass-market customers in a cost-effective manner, including for low-income populations.
- Enhanced efficiency with artificial intelligence and machine learning (AI/ML) taking on a growing portion of investment management responsibilities.
- More customized investment portfolios based on AI/ML targeted customer experiences.
- Development of new return profiles through the use of AI/ML instead of established strategies.

[1] See WEF (2018) for a more detailed discussion.

Box 3. Artificial Intelligence in Credit Underwriting

- Artificial intelligence/machine learning (AI/ML) predictive models can help process credit scoring, enhancing lenders' ability to calculate default and prepayment risks. Research finds that ML reduces banks' losses on delinquent customers by up to 25 percent (Khandani, Adlar, and Lo 2010). There is also evidence that, give their greater accuracy in predicting defaults, automated financial underwriting systems benefit underserved applicants, which results in higher borrower approval rates (Gates, Perry, and Zorn 2002), as does the facilitation of low-cost automated evaluation of small borrowers (Bazarbash 2019).
- AI/ML-assisted underwriting processes enable the harnessing of social, business, location, and internet data, in addition to traditional data used in credit decisions. AI/ML reduces turnaround time and increases the efficiency of lending decisions. Even if a client does not have a credit history, AI/ML can generate a credit score by analyzing the client's digital footprint (social media activity, bills payment history, and search engine activity). AI/ML also has the potential to be used in commercial lending decisions for risk quantification of commercial borrowers.[1] However, financial institutions and supervisors should be cautious in using and assessing AI/ML in credit underwriting and build robust validation and monitoring processes.

[1] See Bazarbash (2019) for a discussion of the potential strengths and weaknesses of AI/ML-based credit assessment.

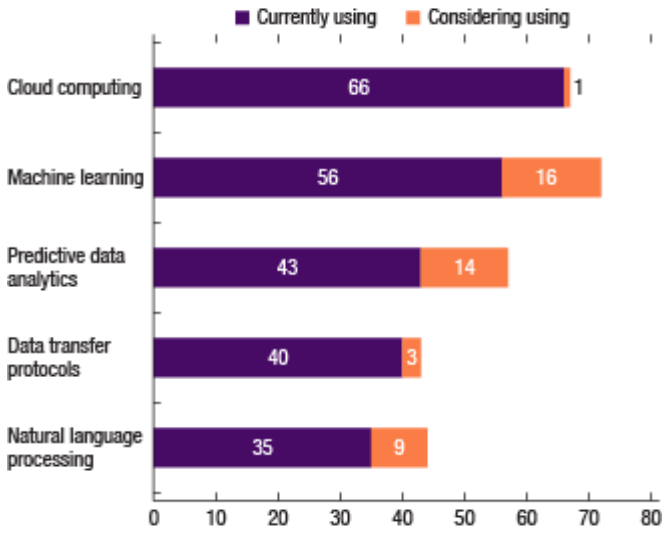
AI/ML introduces new challenges and potential risks. The use of AI/ML in investment and banking depends on the availability of large volumes of good-quality, timely data. With the storage and use of large quantities of sensitive data, data privacy and cybersecurity are of paramount importance. Difficulties in explaining the rationale of AI/ML-based financial decisions is increasingly an important issue as AI/ML algorithms may uncover unknown correlations in data sets that stakeholders may struggle to understand because the underlying causality is unknown. In addition, these models may perform poorly in the event of major and sudden movements in input data resulting in the breakdown of established correlations (for example, in response to a crisis), potentially providing inaccurate decisions, with adverse outcomes for financial institutions or their clients.

C. Risk and Compliance Management

AI/ML advances in recent years are changing the scope and role of technology in regulatory compliance. Regulatory technology (regtech)[6] has assumed greater importance in response to the regulatory tightening and rising compliance costs following the 2008 global financial crisis. For the most part, technology has been used to digitize compliance and reporting processes (Arner, Barberis, and Buckley 2017). However, advances in AI/ML over the past few years are reshaping risk and compliance management by leveraging broad sets of data, often in real time, and automating compliance decisions. This has improved compliance quality and reduced costs.

Maturing AI/ML technology has the potential to propel further adoption of regtech in the financial sector. According to a recent global survey, AI/ML is the top technology under consideration among regtech firms (Schizas and others 2019; Figure 1). Increased adoption of AI/ML in regtech has significantly expanded its use cases, cutting across banking, securities, insurance, and other financial services, and covering a wide variety of activities. These include identity verification, anti-money laundering/combating the financing of terrorism, fraud detection, risk management, stress testing, microprudential and macroprudential reporting, as well as compliance with COVID-19 relief requirements (Box 4).

Figure 1. Top Five Technologies Employed in Regulatory Technology Offerings
(Percent)



Source: Schizas and others (2019).

Regulators have generally been supportive of the adoption of regtech by regulated financial entities. Various regulators (for example, in Hong Kong Special Administrative Region) have developed strategies to promote the adoption of regtech that include boosting awareness, promoting innovation, and enhancing regulatory engagement within the regtech ecosystem. Even where there are no explicit strategies, many authorities have supported regtech adoption. For example, AI/ML systems, in conjunction with facial and voice recognition and NLP capabilities, could play an important role in helping digital banks secure licenses (for example, in Hong Kong Special Administrative Region and Malaysia) and for digital onboarding of customers.

D. Prudential Supervision

Although decisions will ultimately depend on the judgment of supervisors, there is a place for AI/ML, primarily in data collection and data analytics. Many Financial Stability Board member country authorities are currently using ML and NLP tools in data analysis, processing, validation, and plausibility (FSB 2020). With AI, supervisors can draw deeper insights from any type of data and make more informed, data-driven decisions. AI can identify patterns that humans fail to spot and thereby enhance the quality of supervision. It can also make supervision more agile by flagging anomalies to supervisors in real time (ECB 2019). Additionally, supervisory technology (suptech)[7] applications that leverage AI can provide predictive analyses, with the potential to improve the quality of supervision. Despite all its

[6]Regtech refers to the use of technologies by regulated financial entities to meet their regulatory requirements.

[7]Suptech refers to the use of technologies by supervisory agencies to support supervision.

Box 4. Artificial Intelligence in Regulatory Compliance—Sample Use Cases

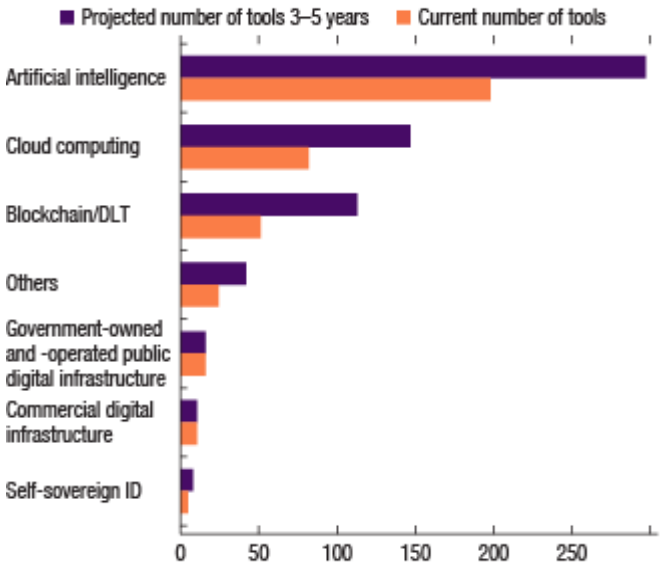
- *Anti-money laundering/combating the financing of terrorism (AML/CFT) compliance.* Artificial intelligence/machine learning (AI/ML)-powered technologies, including analysis of unstructured data and consumer behavior, are used to reduce false positives, which constitute the vast majority of AML/CFT alerts, allowing financial institutions to devote more resources to cases that are likely to be suspect.
- *Mapping and updating regulatory obligations.* AI/ML-based applications are helping financial institutions identify and update relevant regulations, reducing costs and improving regulatory compliance in the process.
- *Conduct risk management.* AI/ML and natural language processing are used to monitor sales calls by financial institutions' employees to ensure compliance with regulatory requirements to accurately disclose the offered financial products' features and risks.
- *Stress-testing.* Many global banks are using AI/ML-enabled data analytics to improve the analysis of complex balance sheets and stress testing models to meet stress testing regulatory requirements.
- *COVID-19 relief.* AI/ML and robotic process automation are used to process COVID-19-mandated credit relief by processing large volumes of applications and loan documents to determine if the loans are eligible for payment modification terms.

potential, AI/ML is not a silver bullet, and the effectiveness of supervision will always very much depend on human judgement and an organization's risk culture.

Currently, suptech use cases are mainly concentrated in misconduct analysis, reporting, and data management (Box 5). Virtual assistance, microprudential, macroprudential, and market surveillance make up a smaller share of cases (di Castri and others 2019). AI in market conduct supervision is used in the collection and analytics of structured and unstructured data,[8] work-flow automation, and conduct risk profiling and early warnings. Off-site surveillance, on-site inspection, and complaints handling, are areas of interest for market conduct authorities (FinCoNet 2020).

Many supervisory authorities are actively exploring the use of AI/ML systems in their risk-based supervision process (Figure 2). AI/ML in microprudential supervision could improve the assessment of risks, including credit and liquidity risks, and governance and risk culture in financial institutions, allowing supervisors to focus on risk analysis and on forward-looking assessments. Risks and trends could be assessed at the sectoral level, improving the quality of macroprudential supervision. AI/ML systems are also applied for market surveillance purposes to detect collusive behavior and price manipulation in the

Figure 2. Technologies Used in Supervisory Technology Tools



Source: FSB (2020).
Note: DLT = distributed ledger technology; "Self-sovereign ID" refers to technologies that give individuals and businesses control of their digital identities.

[8]Unstructured data analytics are used in a number of areas, such as identifying potential violations of advertising guidelines, ascertaining key issues of concern and interest for consumers, and predicting potential harm to consumers in real time using information from social networks.

securities market—potential misconducts that can be especially hard to detect using traditional methods (IIF 2017). The greater use of remote working arrangements as a result of the COVID-19 pandemic is also driving authorities to use technology to improve the effectiveness of remote supervision. Drawing on these experiences, supervisors, post pandemic, might lean more on AI/ML-supported off-site supervision engagements.

Box 5. Artificial Intelligence in Supervision—Sample Applications

Banca d'Italia	- Exploring how loan default forecasting can benefit from the use of machine learning (ML) algorithms using data from different sources.
Banco de España	- Natural language processing (NLP) helps process institutions' environmenta social, and governance disclosures to improve the understanding of the domestic green economy. - Supervised ML assists with misconduct detection.
Bank of Russia	- Operating a system that allows for the analysis of retail loan portfolios using algorithms for processing large data arrays.
Bank of Thailand	- Using an artificial intelligence/ML system to analyze board meeting minutes of financial institutions, which is used by supervisors to assess the regulato compliance of the board and give recommendations as part of the ongoing supervision.
De Nederlandsche Bank	- ML uses transactional data to detect networks of related entities to assess the exposure of financial institutions to networks of suspicious transactions.
European Central Bank	- Machine reading of the "fit and proper" questionnaire helps flag problems. - ML assists in early identification of distress in "less significant institutions." - NLP and ML are used to search for information in supervisory review decisions to facilitate the identification of emerging trends and clusters of risks.
Monetary Authority of Singapore	- Working on a project where credit risk assessments by supervisors are conducted using algorithms instead of sampling.
Oesterreichische Nationalbank	- ML and deep learning algorithms are used to predict the probability that a data set contains errors that need to be rectified by the reporting entity.

The use of AI/ML by supervisory authorities comes with challenges and risks that need to be carefully considered. The effectiveness of AI/ML-driven supervision will depend on data standardization, quality, and completeness, which could be challenging for authorities and regulated institutions alike, particularly when leveraging nontraditional sources of data, such as social media (FSB 2020). Resource and skills gaps could pose a challenge to the effective and safe adoption of AI/ML by supervisory authorities. The personnel pool could be expanded to include AI/ML specialists and data scientists. Finally, deploying AI/ML systems presents supervisors with risks, including those associated with, privacy, cybersecurity, outcome explainability, and embedded bias.

E. Central Banking

AI/ML applications could help central banks implement their mandates, but they have largely been slow to embrace the technology.[9] AI/ML systems could help central banks improve their understanding of economic and financial developments, support better-tuned monetary and macroprudential policies, and improve central banks' operations. They could also strengthen systemic risk monitoring and potentially help predict the buildup of systemic risks and speed up crisis response.[10] Notwithstanding these potential benefits, the use of AI/ML processes for policymaking should remain subject to good judgment. Furthermore, AI/ML could provide increased efficiency and better internal control opportunities for central banks, including monitoring of internal operations and resource allocation across functions.[11] The technology to develop these applications largely exists; central banks' slow adoption thereof can be attributed to cultural, political, and legal factors, and lack of adequate capacity (Danielsson, Macrae, and Uthemann 2020). Nevertheless, some central banks are exploring the possibilities of AI/ML (Box 6). Experiments and research by central banks have focused on improving the near-term forecasting capacity and the monitoring of market sentiment to inform policy decisions. Some central banks are also developing applications to improve internal processes and back-office functions (for example, cash management).

AI/ML use in central banking does not seem to raise significant concerns for the time being, but this could change with broader deployment. Expanded use of large nontraditional and unstructured data sets to produce analytics in support of monetary and macroprudential policymaking could expose a central bank to data biases that are usually avoided in sampling methods. ML algorithms are still susceptible to sudden structural shifts in the data from unforeseen events, and the resulting errors could undermine the effectiveness of a central bank's crisis monitoring and response capacity. Central banks could also face challenges in acquiring quality and representative data as well as data privacy and security issues.[12] These concerns would be amplified if the central bank does not have the resources and skills required to safely operate AI/ML or mitigate the risks associated with it.

[9]The focus of this section is on central banks' monetary and macroprudential functions. AI/ML could also be relevant for other, potential central bank functions, in such areas as consumer protection, financial integrity, financial inclusion, or even climate change.

[10]Conceivably, AI/ML could strengthen the central bank's rapid crisis response to a fast-moving financial crisis.

[11]See Khan and Malaika (2021) for more details.

[12]For a more detailed discussion, see Doerr, Gambacorta, and Serena (2021).

Box 6. Artificial Intelligence in Central Banking—Sample Applications

Strengthening nowcasting

- Sveriges Riksbank has developed real-time indicators to support its policy analysis. These include investigating whether fruit and vegetable prices, scraped daily from the Internet and machine learning (ML)-processed into an index, could improve the accuracy of short-term inflation forecasts.
- Reserve Bank of New Zealand is experimenting with using ML to process large, real-time data sets of about 550 macroeconomic indicators to improve its nowcasts of GDP growth, with results so far outperforming comparable statistical benchmarks.

Assessing market sentiment

- Banca d'Italia has developed a real-time tracking system of consumer inflation expectations, using ML and textual analysis of millions of daily Italian Twitter feeds. The Twitter-based indicators seem to be highly correlated with standard statistical measures, provide superior forecasting for survey-based monthly inflation expectations than all other available sources, and accurately anticipate consumers' expectations.
- Bank Indonesia is testing ML-based techniques for identifying the expectation of stakeholders, scraped from published news articles, of Bank Indonesia's policy rate, in support of the monthly Board of Governor's meeting.

Monitoring uncertainty

- Banco de Mexico staff has built a sentiment-based risk index using an artificial intelligence (AI)/ML system to analyze Twitter messages in response to positive or negative shocks to the Mexican financial sector. The research finds that index shocks correlate positively with an increase in financial market risk, stock market, volatility, sovereign risk, and foreign exchange rate volatility.
- Banco Central de Chile staff has developed a daily-frequency index of economic uncertainty for Chile using AI/ML to analyze Twitter feeds. It tracks the level of general disagreement—a proxy for economic uncertainty—on economic developments and policies. The index shows significant spikes that coincide with several episodes of substantial economic uncertainty of both local and international origin.
- De Nederlandsche Bank is exploring whether AI is useful for detecting liquidity problems at banks in anticipation of potential deposit runs (Triepels, Daniels, and Heijmans 2018).

Improving internal processes

- Banco de España has developed an AI tool for sorting banknotes between fit and unfit for circulation.

3. Risks and Policy Considerations

The rapid deployment of AI/ML systems in finance will have significant impact that will require robust policy responses to ensure the integrity and safety of the financial system. Concerns are rising regarding a number of issues, such as embedded bias in AI/ML systems, the ability to explain the rationale of their decisions, their robustness (particularly with respect to cyber threats and privacy), and their potential impact on financial stability. This chapter discusses those concerns. Annex 2 provides a consolidated profile of risks arising from AI/ML use in the financial sector.

A. Embedded Bias

The growing use of AI/ML in the financial sector, which is highly regulated and where public trust is an essential factor, has stirred discussions on the risk of embedded bias. Friedman and Nissenbaum (1996) defines embedded bias as computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others.[13] Customer categorization processes used in AI/ML can lead to bias in the financial sector through differentiation in pricing or service quality. Bias in AI/ML decisions is often the result of biased training data that comes from existing biased processes and data sets, which will teach AI/ML models to be biased too (Wang 2016). Data biases, or inaccurate and insufficient information, could result in greater financial exclusion and feed distrust for the technology, especially among the most vulnerable (Sahay and others 2020). Data collection could result in bias in two ways:

- The data used to train the system might be incomplete or unrepresentative. For example, predictive algorithms (such as for a loan approval) favor groups that are better represented in the training data, given that there will be less uncertainty associated with those predictions (Goodman and Flaxman 2016).
- The data may underpin prevailing prejudices (Hao 2019). For example, Amazon discovered that its internal recruiting tool was dismissing female candidates because it was trained on historical hiring decisions, which favored men over women.[14]

Human bias can give rise to bias in the algorithm during the design and training of AI/ML systems. For instance, when a researcher decides on the features to include or exclude in the ML model, the choice can be influenced by various psychological, social, emotional, and cultural factors, which in turn influence the researcher's feedback on the outputs when training the AI/ML system.[15]

Notwithstanding the potential for bias in AI/ML systems, they may help reduce prevailing biases. Mayson (2019) and Silberg and Manyika (2019) argue that AI/ML systems can reduce human bias in decision making because ML algorithms can eliminate irrational biases that stem from the subjective interpretation of data.[16] Miller (2018) notes that AI systems, while susceptible to embedded bias, can still improve the decision-making process by mitigating human bias. Finally, Silberg and Manyika (2019) suggests that, even though many AI/ML systems could be perceived as black boxes, their prediction and decision-making processes could be scrutinized more than those of individuals and, thus, existing biases could be identified and mitigated.

[13]This bias should not be confused with statistical bias, which is defined as the difference between the expected value of the estimator and its true value (Lehmann 1951). For further discussion on bias-variance trade-off, see Annex 1.
[14]The algorithm was taught to recognize word patterns; AI software penalized any resume that contained the word "women."
[15]For an in-depth discussion of how psychological, social, emotional, and cultural factors already play a role in the financial sector, see Khan (2018).
[16]For an in-depth discussion on the relation between human bias and algorithms, see Plous (2002); Corbett-Davies and others (2017); and Kleinberg and others (2018a, 2018b, and 2019).

Given that bias may arise as an unintended consequence of AI/ML systems, regulators could view this as a potential source of operational and reputational risks. Financial institutions that deploy AI/ML systems in a significant manner, particularly with regard to the provision of credit and financial services and in risk management, should develop and implement bias mitigation and detection plans as part of their operational risk management framework. Such plans could include adequate assurances about the algorithm's robustness against systemically generating biased decisions, disclosure of data sources, awareness of potential bias-generating factors in the data, monitoring and evaluation tools, and, more generally, how the institution intends to meet the standing anti-discrimination rules in the context of AI/ML deployment.

Policy responses to AI/ML embedded bias issues could be advanced by developing and deploying a broader framework for the governance and ethical use of AI/ML applications. In recent years, several initiatives have been launched to help develop such a framework. For example, in April 2019 the European Union released its "Ethics Guidelines for Trustworthy AI," which outlines key requirements for a trustworthy AI system.[17] Similarly, in May 2019 the Organisation for Economic Co-operation and Development adopted a set of principles to promote ethical and innovative AI development. These principles contributed to the June 2019 Group of Twenty declaration on "human-centered" AI.

B. Unboxing the "Black Box": Explainability and Complexity

Explainability of AI/ML systems outcomes is an important issue, particularly when used in the financial sector. ML models are often referred to as black boxes because they are not directly explainable by the user (Guidotti and others 2019). This characteristic could make detection of the appropriateness of ML decisions difficult[18] and could expose organizations to vulnerabilities—such as biased data, unsuitable modeling techniques, or incorrect decision making—and potentially undermine the trust in their robustness (Silberg and Manyika 2019).

Explainability is a complex and multifaceted issue. There are several reasons why ML models are frequently considered to be black boxes: (1) they are complicated and cannot be easily interpreted, (2) their input signals might not be known, and (3) they are an ensemble of models rather than a single independent model. Furthermore, stronger explainability might enable outsiders to manipulate the algorithm (Molnar 2021) and create risks in the financial system. Generally, there is a trade-off between model flexibility—which refers to its capacity to approximate different functions and is directly related to the number of parameters of the model—and its explainability. ML models are more flexible and accurate, but are less explainable than, for example, linear models that produce easily interpretable results but are less accurate. Box 7 provides a brief overview of ML explanation methods.

Ongoing research and regulatory initiatives may help address AI/ML explainability issues. The literature points to various levels of explainability related to different stages of the modeling process, or different objectives, such as explaining individual predictions or the overall model behavior (Guidotti and others 2019). To advance the policy discussion, regulators could explore the possibility that financial sector AI/ML models may require different levels of explainability, depending on the impact of the model's outcome or the governing regulation. Broadly, regulatory guidance that creates a proper framework and strategy for managing explainability risks at different levels is needed for the financial sector. Currently, ML explainability is included in only a few regulatory frameworks as part of the

[17]Similarly, the Monetary Authority of Singapore published "Principles to Promote Fairness, Ethics, Accountability, and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector" in November 2018; De Nederlandsche Bank published "General Principles for Use of Artificial Intelligence in Finance" in July 2019; and the Hong Kong Monetary Authority published "High-Level Principles on Artificial Intelligence" in November 2019).

[18]In this context, "explainability" (sometimes referred to as "interpretability") in a model means showing how input variables contribute to both the model's aggregate results and explain individual outcomes (HKMA 2020).

accountability and transparency requirements (for example, in the European Union, Hong Kong Special Administrative Region, Netherlands, and Singapore).

Box 7. Explaining the "Black Box"

- *Explainability by design.* Some machine learning (ML) models are inherently interpretable by design, such as linear models and decision trees, but this but may constrain their accuracy.
- *Model inspection.* During model development, various techniques could be used for testing the sensitivity of a model to changes in variables. This approach can help avoid unintended bias and other modeling pitfalls, however, the techniques have limited applicability to deployed models.
- *Individual prediction explanation.* After the model is deployed, some methods can provide local approximations that help explain individual predictions.[1] For instance, why credit was or was not granted to an individual or why a buy or sell order was posted, thereby making it easier to audit results.
- *Global model explanation.* This method consists of using an interpretable model, such as linear regression or decision tree, to explain the behavior of the more complex ML models used for prediction. The interpretable model would not be used for prediction, but to provide a general explanation of the behavior of the main model. While useful to comprehend the overall behavior of the predictive model, there is no assurance that the surrogate model can fully mimic the more complex model while keeping a reasonable level of explainability.
- *Models to explain ML models*. Research is increasingly focused on developing techniques that use ML models or their outcomes as training input to achieve explainability. This is accomplished by repeatedly disturbing the ML inputs or variables and comparing them to the ML outcome to ascertain the decision-making process.[2]

[1]For example, Shapley values (Shapley 1953) and Lime values (Ribeiro, Singh, Guestrin 2016).
[2]See examples of explainability in Cloudera Fast Forward (2020).

C. Cybersecurity

AI/ML adoption increases the scope for cyber threats and brings in new unique cyber risks. In addition to traditional cyber threats from human or software failures, AI/ML systems are vulnerable to novel threats. Such threats focus on manipulating data at some stage of the AI/ML lifecycle to exploit inherent limitations of AI/ML algorithms (Comiter 2019). Such manipulation allows attackers to evade detection and prompts AI/ML to make the wrong decisions or to extract information. Due to their complexity and the potential impact for financial sector institutions, ML models require constant oversight to certify that such attacks are detected consistently and managed promptly. In addition to typical cybersecurity concerns, specific cyber threats to AI/ML can broadly be grouped as:

- Data poisoning attacks intend to influence an ML algorithm during the training stage by adding special samples to its training data set. These attacks cause the AI to incorrectly learn to classify or recognize information. Data poisoning also may be used to create Trojan models, which hide malicious actions that wait for special inputs to be activated (Liu, Dolan-Gavitt, and Garg 2018). Data poisoning attacks require privileged access to model and training information, but once executed successfully, and as long as the malicious behavior does not interfere in regular diagnostics tests, infected models may be undetectable.

- Input attacks[19] allow attackers to introduce perturbations to data inputs and mislead AI systems during operations. For example, attackers could alter images with elements unperceivable to human vision, but which provoke AI/ML image recognition systems to mislabel the images.
- Model extraction or model inversion attacks attempt to recover training input and data, or the model itself. Membership inference (a variant) simply aims to check whether a specific data instance was in the training set. Such attacks may be performed as black-box attacks, whereby attackers have merely read-only access to the model, possibly through business application programming interfaces. Current privacy protection laws (for example, the European Union's General Data Protection Regulation) are not designed to deal with the methods of model inversion attacks, raising privacy and copyright concerns.

AI/ML cybersecurity is increasingly becoming a concern for financial sector regulators. AI/ML cyber threats could undermine the integrity of, and trust in, the financial sector. Corrupted systems could undermine the financial sector's capacity to accurately assess, price, and manage risks, which could lead to the buildup of unobserved systemic risks. Attackers could also acquire training data sets that contain sensitive financial and personal information.

The regulatory perimeter of cybersecurity requirements in the financial sector could be expanded to cover AI/ML-specific cyber threats. Providers and users of AI/ML applications in the financial sector should be required to put in place, as part of their broader cybersecurity framework, mitigating strategies. These could include detection and reporting systems, robust protection of training data feeds, and strategies to secure model and data privacy.

D. Data Privacy

AI/ML introduces new unique privacy issues. Privacy concerns regarding big data are well known and predate the emergence of AI/ML into the mainstream.[20] Tools have been developed to help maintain data anonymity and data subjects' privacy. Legal data policy frameworks are being put in place across the globe to address these concerns. The robustness of AI/ML models, however, in preventing data leakage from the training data set raises new privacy concerns. For example, AI/ML has the capacity to unmask anonymized data through inferences (that is, deducing identities from behavioral patterns). Similarly, AI/ML may remember information about individuals in the training set after the data is used, or AI/ML's outcome may leak sensitive data directly or by inference. Tools are being developed to address these issues and strengthen AI/ML models' robustness in safeguarding sensitive data, but more work is needed, along with an appropriate update to the legal and regulatory framework that requires AI/ML systems and related data sources to adhere to enhanced privacy standards, along with relevant anti-money laundering/combating the financing of terrorism requirements.

E. Robustness

Robust AI/ML algorithms will help build public trust in an AI-driven financial system and safeguard financial stability and the integrity of the financial system. This is particularly important, given the relatively high concentration of AI/ML service providers and the lack of adequate oversight capacity in many jurisdictions to effectively engage financial institutions using AI/ML systems. Several challenges would need to be addressed by the industry and regulators to ensure the robustness of ML models with respect to cyber threats and safeguarding privacy, as discussed above, as well as with respect to their performance. The latter covers issues related to minimizing ML models' false signals

[19]Input attacks are sometimes referred in specialized literature as "adversarial examples" (a form of adversarial attack). Furthermore, "input attack" should not be confused with "input validation attack," a popular vulnerability often used to hack websites.

[20]See Haksar and others (2021) on broader data policy issues.

during periods of structural shifts and having in place an appropriate governance over the AI/ML development process.

AI/ML systems in the financial sector have performed well in a data environment that is relatively stable and produces reliable signals, but that could quickly change in periods of rapid structural shift. In a relatively stable environment, AI/ML models are reasonably able to incorporate evolving data trends without significant loss in prediction accuracy. However, they face a more challenging task with structural changes in their data environment, when a previously reliable signal becomes unreliable or when behavioral correlations shift significantly. The recently observed misalignment of AI/ML-generated risk assessments during the COVID-19 pandemic is a good illustration. The crisis negatively affected the performance of ML models because they were not originally trained for such an event. Harker (2020) and BoE (2020) highlight that ML algorithms trained using pre–COVID-19 data may experience performance deterioration, for example, by pointing to aggregate credit scores in the United States, which, despite the record job loss and rising defaults, have improved during the acute part of the crisis, reflecting temporary relief measures that were not captured by the algorithms.

New governance frameworks for the development of AI/ML systems are needed to strengthen prudential oversight and to avoid unintended consequences. Given that the process of creating AI/ML systems is very similar to software development, proper quality control and agility of the process is required. The process could encompass all phases of development, testing, and deployment, and focus on the relevant risks and controls. Checks for embedded bias, data poisoning, other security risks, and performance should comply with best practices of software development, including separation of duties and constant monitoring.

F. Impact on Financial Stability

The widespread deployment of AI/ML systems in the financial sector will be transformational, and their impact on financial stability is yet to be fully assessed. As highlighted above, on one hand, with carefully designed and tested algorithms satisfying a high level of controls to limit risks and performance issues, AI/ML systems may bring increased efficiencies; better assessment, management, and pricing of risks; improved regulatory compliance; and new tools for prudential surveillance and enforcement—all of which will contribute positively to financial stability. On the other hand, AI/ML systems bring new and unique risks arising from the opacity of their decisions, susceptibility to manipulation, robustness issues, and privacy concerns. These could undermine the public's trust in the integrity and safety of an AI/ML-driven financial system. Furthermore, AI/ML could potentially bring about new sources and transmission channels of systemic risks. More specifically:

- AI/ML service providers could become systemically important participants in financial market infrastructure due to the high specialization of AI/ML systems and network effects, which could increase the financial system's vulnerability to single points of failure.
- The concentration of third-party AI/ML algorithm providers could drive greater homogeneity in risk assessments and credit decisions in the financial sector, which, coupled with rising interconnectedness, could create the conditions for a buildup of systemic risks. The likely concentration of data and growing use of alternative data in AI/ML could result in, respectively, the risk of uniformity (herding) and out-of-sample risk that eventually could lead to systemic risk.
- The widespread use of AI/ML could potentially increase procyclicality of financial conditions. For instance, credit underwriting and risk management processes are inherently procyclical, reacting to financial conditions in ways that reinforce and amplify those changes (that is, falling default rates lead to more lending, which can further lower default rates). AI/ML may automate and accelerate the procyclicality and potentially obscure it (due to explainability issues).

- Moreover, in the case of a tail risk event, an inaccurate risk assessment and reaction by ML algorithms could quickly amplify and spread the shock throughout the financial system and complicate or even undermine the effectiveness of the policy response.
- Challenges in interpretation, sustainability of analytical power, and prediction of models raise the concern that economic policies or market strategies based on these models will be challenging to interpret or predict by the relevant counterparties, creating additional asymmetric information in the market, with an uncertain impact on financial stability.
- Finally, regulatory gaps could adversely impact financial stability if technological advances outpace existing regulations. Such advances are often led by providers who may fall outside existing regulatory perimeters.[21]

The rapid evolution of AI/ML has led to a range of regulatory responses. While some jurisdictions have taken a more holistic approach to addressing the issues involved (for example, the Monetary Authority of Singapore[22] and De Nederlandsche Bank[23]), others have concluded that existing regulations and expectations on good governance are sufficient to address the emerging issues. Whether it is by way of new regulations or existing ones, regulators have generally focused on AI/ML governance frameworks, risk management, internal controls, and better controls over the model and data.

Addressing these challenges requires broad regulatory and collaborative efforts. An adequate policy response requires developing clear minimum standards and guidelines for the sector, coupled with stronger focus on securing the necessary technical skills. Given the inherent interconnectivity of issues related to the deployment of AI/ML systems in the financial sector, collaboration among financial institutions, central banks, financial supervisors, and other stakeholders is important to avoid duplication of work and to help counter potential risks. Many leading jurisdictions in the AI/ML sector have relied on well-articulated national AI strategies for promoting AI/ML development while ensuring that regulatory gaps do not materialize. Annex 3 provides an overview of country approaches to developing AI national strategies.

[21]For a more detailed discussion of new transmission channels of systemic risks, see Gensler and Bailey (2020).
[22]See MAS (2018).
[23]See DNB (2019).

4. Conclusion

The deployment of artificial intelligence (AI) and machine learning (ML) systems in the financial sector will continue to accelerate. This trend is driven by rapid increases in computational powers, data storage capacity, and big data, as well as by significant progress in modeling and use-case adaptations. The COVID-19 pandemic is accelerating the shift toward a more contactless environment and increasingly digital financial services, which will further strengthen the appeal of AI/ML systems to providers of financial services.

Use of AI/ML will bring important benefits but will also raise significant financial policy challenges. AI/ML systems offer financial institutions the potential for significant cost savings and efficiency gains, new markets, and better risk management; bring customers new experiences, products, and lower costs; and offer powerful tools for regulatory compliance and prudential oversight. However, these systems also bring about ethical questions and new unique risks to the financial system's integrity and safety, of which the full extent is yet to be assessed. The task facing financial sector policymakers is further complicated by the fact that these innovations are still evolving and morphing as new technologies come into play. These developments call for improvements in oversight monitoring frameworks and active engagement with stakeholders to identify possible risks and remedial regulatory actions.

In step with the Bali Fintech Agenda's call on national authorities to embrace the fintech revolution, regulators should broadly welcome the advancements of AI/ML in finance and undertake the preparations to capture its potential benefits and mitigate its risks. This includes the timely strengthening of institutional capacity, recruiting relevant expertise, building up knowledge, improving external communication with stakeholders, and expanding consumer education. Deployment of AI/ML systems in the financial sector has proven to be most effective when there are national AI strategies in place that involve all relevant public and private bodies.

Cooperation and knowledge sharing at the regional and international level is becoming increasingly important. This would allow for the coordination of actions to support the safe deployment of AI/ML systems and the sharing of experiences and knowledge. Cooperation will be particularly important to ensure that less-developed economies have access to knowledge related to techniques and methods, use cases, and regulatory and supervisory approaches.

Finally, the evolving nature of the AI/ML technology and its applications in finance mean that neither the users, the technology providers and developers, nor the regulators understand, currently, the full extent of the strengths and weaknesses of the technology. Hence, there may be many unexpected pitfalls that are yet to materialize, and countries will need to strengthen their monitoring and prudential oversight.