Artificial Intelligence, Machine Learning and Big Data in Finance

Opportunities, Challenges and Implications for Policy Makers

OECD

Artificial Intelligence, Machine Learning and Big Data in Finance

Opportunities, Challenges, and Implications for Policy Makers

Please cite this publication as:

OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*, https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm.

Foreword

Artificial Intelligence (AI) techniques are being increasingly deployed in finance, in areas such as asset management, algorithmic trading, credit underwriting or blockchain-based finance, enabled by the abundance of available data and by affordable computing capacity. Machine learning (ML) models use big data to learn and improve predictability and performance automatically through experience and data, without being programmed to do so by humans.

The deployment of AI in finance is expected to increasingly drive competitive advantages for financial firms, by improving their efficiency through cost reduction and productivity enhancement, as well as by enhancing the quality of services and products offered to consumers. In turn, these competitive advantages can benefit financial consumers by providing increased quality and personalised products, unlocking insights from data to inform investment strategies and potentially enhancing financial inclusion by allowing for the analysis of creditworthiness of clients with limited credit history (e.g. thin file SMEs).

At the same time, AI applications in finance may create or intensify financial and non-financial risks, and give rise to potential financial consumer and investor protection considerations (e.g. as risks of biased, unfair or discriminatory consumer results, or data management and usage concerns). The lack of explainability of AI model processes could give rise to potential pro-cyclicality and systemic risk in the markets, and could create possible incompatibilities with existing financial supervision and internal governance frameworks, possibly challenging the technology-neutral approach to policymaking. While many of the potential risks associated with AI in finance are not unique to this innovation, the use of such techniques could amplify these vulnerabilities given the extent of complexity of the techniques employed, their dynamic adaptability and their level of autonomy.

The report can help policy makers to assess the implications of these new technologies and to identify the benefits and risks related to their use. It suggests policy responses that that are intended to support AI innovation in finance while ensuring that its use is consistent with promoting financial stability, market integrity and competition, while protecting financial consumers. Emerging risks from the deployment of AI techniques need to be identified and mitigated to support and promote the use of responsible AI. Existing regulatory and supervisory requirements may need to be clarified and sometimes adjusted, as appropriate, to address some of the perceived incompatibilities of existing arrangements with AI applications.

Acknowledgements

This report has been prepared by *Iota Kaousar Nassr* under the supervision of *Robert Patalano* from the Division of Financial Markets of the OECD Directorate for Financial and Enterprise Affairs. *Pamela Duffin* and *Ed Smiley* provided editorial and communication support.

The report supports the work of the OECD Committee on Financial Markets and its Experts Group on Finance and Digitalisation. It was discussed by the Committee in April 2021, and is a product of the Committee's Expert Group on Finance and Digitalisation.

The author gratefully acknowledges valuable input and constructive feedback provided by the following individuals and organisations: *Anne Choné*, ESMA; *Nancy Doyle*, US Commodity Futures Trading Commission; *Adam Głogowski* and *Paweł Pisany*, National Bank of Poland; *Peter Grills*, US Treasury; *Alex Ivančo*, Ministry of Finance of the Czech Republic; *Antonina Levashenko* and *Ivan Ermokhin*, Russia OECD Centre RANEPA; *Aleksander Madry*, MIT; *Irina Mnohoghitnei* and *Mohammed Gharbawi*, Bank of England; *Benjamin Müller*, Swiss National Bank; *Borut Poljšak*, Bank of Slovenia; *Merav Shemesh* and *Itamar Caspi*, Bank of Israel; *Akiko Shintani*, Permanent Delegation of Japan to the OECD and *Yuta Takanashi*, *Ryosuke Ushida*, and *Ayako Yamazaki*, Financial Services Agency, Japan; *Ilaria Supino, Giuseppe Ferrero*, *Paola Masi* and *Sabina Marchetti,* Banca d'Italia. The report has also benefited from views and input provided by academia and the industry.

This report contributes to the horizontal OECD Going Digital project which provides policy makers with tools to help economies and societies prosper in an increasingly digital and data-driven world. For more information, visit www.oecd.org/going-digital.

Table of contents

FIGURES

INFOGRAPHICS

Executive Summary

Artificial intelligence (AI) in finance

Artificial intelligence (AI) systems are machine-based systems with varying levels of autonomy that can, for a given set of human-defined objectives, make predictions, recommendations or decisions. AI techniques are increasingly using massive amounts of alternative data sources and data analytics referred to as '**big data**'. Such data feed **machine learning (ML) models which use such data** to learn and improve predictability and performance automatically through experience and data, without being programmed to do so by humans.

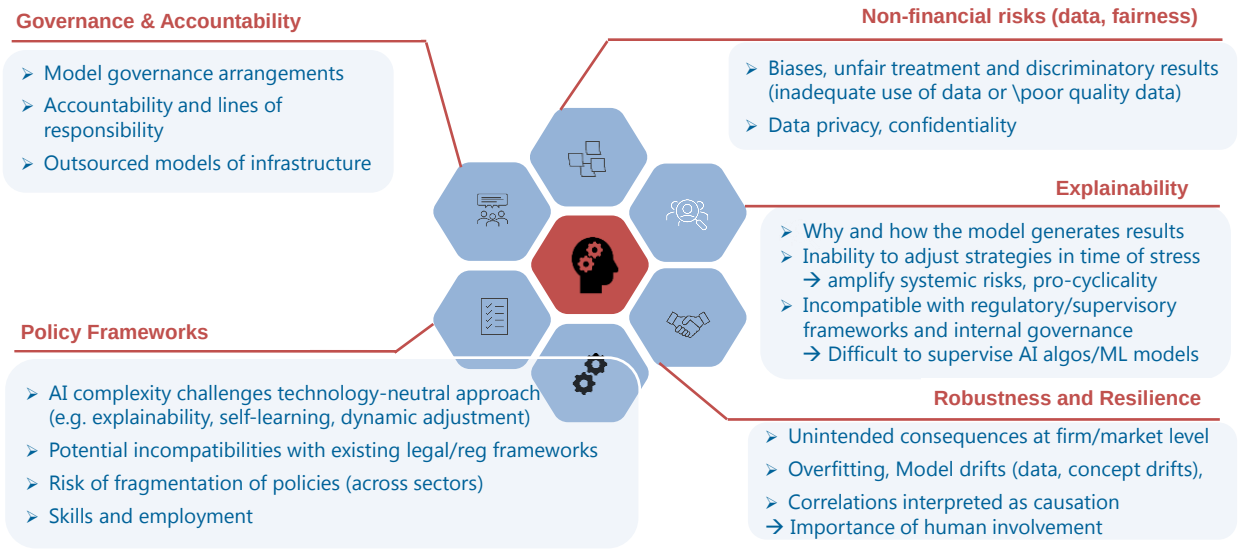
The COVID-19 crisis has accelerated and intensified the digitalisation trend that was already observed prior to the pandemic, including around the use of AI. Global spending on AI is forecast to double over the period 2020-24, growing from USD50 bn in 2020 to more than USD110 bn in 2024 (IDC, 2020[1]). Growing AI adoption in finance, in areas such as asset management, algorithmic trading, credit underwriting or blockchain-based financial services, is enabled by the abundance of available data and by increased, and more affordable, computing capacity.

The deployment of AI in finance is expected to increasingly drive competitive advantages for financial firms, through two main avenues: (a) by improving the firms' efficiency through cost reduction and productivity enhancement, therefore driving higher profitability (e.g. enhanced decision-making processes, automated execution, gains from improvements in risk management and regulatory compliance, back-office and other process optimisation); and (b) by enhancing the quality of financial services and products offered to consumers (e.g. new product offering, high customisation of products and services). Such competitive advantage can, in turn, benefit financial consumers, either through increased quality of products, variety of options and personalisation, or by reducing their cost.

Why is the deployment of AI in finance relevant to policy makers

AI applications in finance may create or intensify financial and non-financial risks, and give rise to potential financial consumer and investor protection considerations. The use of AI amplifies risks that could affect a financial institution's safety and soundness, given the lack of explainability or interpretability of AI model processes, with potential for pro-cyclicality and systemic risk in the markets. The difficulty in understanding how the model generates results could create possible incompatibilities with existing financial supervision and internal governance frameworks, while it may even challenge the technology-neutral approach to policymaking. AI may present particular risks of consumer protection, such as risks of biased, unfair or discriminatory consumer results, or data management and usage concerns. While many of the potential risks associated with AI in finance are not unique to AI, the use of AI could amplify such vulnerabilities given the extent of complexity of the techniques employed, the dynamic adaptability of AI-based models and their level of autonomy for the most advanced AI applications.

Figure 1. Relevant issues and risks stemming from the deployment of AI in finance



Source: OECD staff illustration.

How is AI affecting parts of the financial markets?

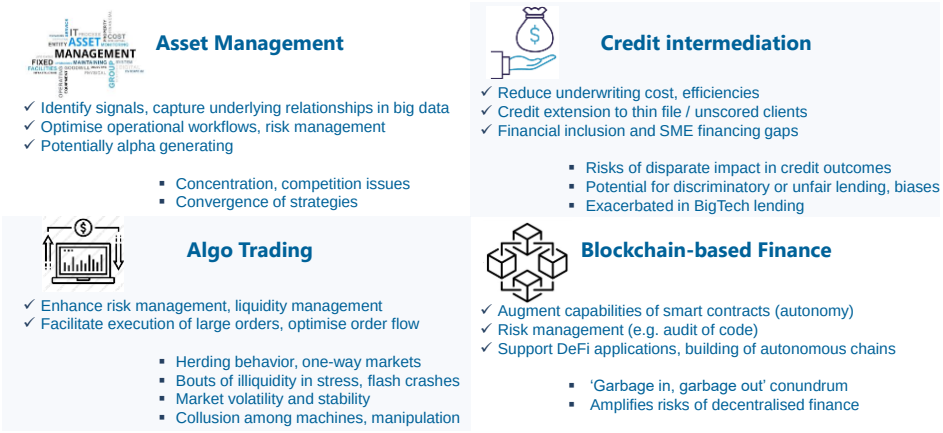
AI techniques are applied in asset management and the buy-side activity of the market for asset allocation and stock selection based on ML models' ability to identify signals and capture underlying relationships in big data, as well as for the optimisation of operational workflows and risk management. The use of AI techniques may be reserved to larger asset managers or institutional investors who have the capacity and resources to invest in such technologies.

When used in trading, AI adds a layer of complexity to conventional algorithmic trading, as the algorithms learn from data inputs and dynamically evolve into computer-programmed algos, able to identify and execute trades without any human intervention. In highly digitised markets, such as equities and FX markets, AI algorithms can enhance liquidity management and execution of large orders with minimal market impact, by optimising size, duration and order size in a dynamic fashion, based on market conditions. Traders can also deploy AI for risk management and order flow management purposes to streamline execution and produce efficiencies.

Similar to non-AI models and algos, the use of the same ML models by a large number of finance practitioners could potentially prompt of herding behaviour and one-way markets, which in turn may raise risks for liquidity and stability of the system, particularly in times of stress. Although AI algo trading can increase liquidity during normal times, it can also lead to convergence and by consequence to bouts of illiquidity during times of stress and to flash crashes. Market volatility could increase through large sales or purchases executed simultaneously, giving rise to new sources of vulnerabilities. Convergence of trading strategies creates the risk of self-reinforcing feedback loops that can, in turn, trigger sharp price moves. Such convergence also increases the risk of cyber-attacks, as it becomes easier for cyber-criminals to influence agents acting in the same way. The abovementioned risks exist in all kinds of algorithmic trading, however, the use of AI amplifies associated risks given their ability to learn and dynamically adjust to evolving conditions in a fully autonomous way. For example, AI models can identify signals and learn the impact of herding, adjusting their behaviour and learning to front run based on the earliest of signals. The scale of complexity and difficulty in explaining and reproducing the decision mechanism of AI algos and models makes it challenging to mitigate these risks.

AI techniques could exacerbate illegal practices in trading aiming to manipulate the markets, and make it more difficult for supervisors to identify such practices if collusion among machines is in place. This is enabled due to the dynamic adaptive capacity of self-learning and deep learning AI models, as they can recognise mutual interdependencies and adapt to the behaviour and actions of other market participants or other AI models, possibly reaching a collusive outcome without any human intervention and perhaps without the user even being aware of it.

Figure 2. Impact of AI on business models and activity in the financial sector



Source: OECD Staff.

AI models in lending could reduce the cost of credit underwriting and facilitate the extension of credit to 'thin file' clients, potentially promoting financial inclusion. The use of AI can create efficiencies in data processing for the assessment of creditworthiness of prospective borrowers, enhance the underwriting decision-making process and improve the lending portfolio management. It can also allow for the provision of credit ratings to 'unscored' clients with limited credit history, supporting the financing of the real economy (SMEs) and potentially promoting financial inclusion of underbanked populations.

Despite their vast potential, AI-based models and the use of inadequate data (e.g. relating to gender or race) in lending can raise risks of disparate impact in credit outcomes and the potential for biased, discriminatory or unfair lending. In addition to inadvertently generating or perpetuating biases, AI-driven models make discrimination in credit allocation even harder to find, and outputs of the model difficult to interpret and communicate to declined prospective borrowers. Such challenges are exacerbated in credit extended by BigTech that leverage their access to vast sets of customer data, raising questions about possible anti-competitive behaviours and market concentration in the technology aspect of the service provision (e.g. cloud).

The use of AI techniques in blockchain-based finance could enhance the potential efficiency gains in DLT-based systems and augment the capabilities of smart contracts. AI can increase the autonomy of smart contracts, allowing the underlying code to be dynamically adjusted according to market conditions. The use of AI in DLT systems also introduces, if not amplifies, challenges encountered in AI-based traditional financial products, such as lack of interpretability of AI decision-making mechanisms and difficulty in supervising networks and systems based on opaque AI models. At the moment, AI is mostly being used for risk management of smart contracts, for the identification of flaws in the code. It should be noted, however, that smart contracts have existed long before the advent of AI applications and rely on simple software code. As of today, most smart contracts used in a material way do not have ties to AI techniques and many of the suggested benefits from the use of AI in DLT systems remains theoretical at this stage.

In the future, AI could support decentralised applications in decentralised finance ('DeFi'), by enabling automated credit scoring based on users' online data, investment advisory services and trading based on financial data, or insurance underwriting. In theory, AI-based smart contracts that are self-learned[1] and adjust dynamically without human intervention could result in the building of fully autonomous chains. The use of AI could promote further disintermediation by replacing off-chain third-party providers of information with AI inference directly on-chain. It should be noted, however, that AI-based systems do not necessarily resolve the garbage in, garbage out conundrum and the problem of poor quality or inadequate data inputs observed in in blockchain-based systems. This, in turn, gives rise to significant risks for investors, market integrity and the stability of the system, depending on the size of the DeFi market. Equally, AI could amplify the numerous risks experienced in DeFi markets, adding complexity to already hard-to-supervise autonomous DeFi networks without single regulatory access points or governance frameworks that allow for accountability and compliance with oversight frameworks.

Key overriding risks and challenges, and possible mitigating actions

The deployment of AI in finance could amplify risks already present in financial markets given their ability to learn and dynamically adjust to evolving conditions in a fully autonomous way, and give rise to new overriding challenges and risks. Existing risks are associated with the inadequate use of data or the use of poor quality data that could allow for biases and discriminatory results, ultimately harming financial consumers. Concentration risks and related competition issues could result from the investment requirements of AI techniques, which could lead to dependence on a few large players. Market integrity and compliance risks could stem from the absence of adequate model governance that takes into account the particular nature of AI, and from the lack of clear accountability frameworks. Risks are also associated with oversight and supervisory mechanisms that may need to be adjusted for this new technology. Novel risks emerging from the use of AI relate to the unintended consequences of AI-based models and systems for market stability and market integrity. Important risks stem from the difficulty in understanding how AI-based models generate results (explainability). Increased use of AI in finance could lead to potential increased interconnectedness in the markets, while a number of operational risks related to such techniques could pose threat to the resilience of the financial system in times of stress.

The use of big data in AI-powered applications could introduce an important source of non-financial risk driven by challenges and risks related to the quality of the data used; data privacy and confidentiality; cyber security; and fairness considerations. Depending on how they are used, AI methods have the potential to help avoid discrimination based on human interactions, or intensify biases, unfair treatment and discrimination in financial services. Biases and discrimination in AI can result from the use of poor quality, flawed or inadequate data in ML models, or unintentionally through inference and proxies (for example, inferring gender by looking into purchasing activity data). In addition to financial consumer protection considerations, there are potential competition issues arising from the use of big data and ML models, relating to high concentration amongst market providers in some markets or increased risks of tacit collusions.

The most widely acknowledged challenge of ML models is the difficulty in understanding why and how the model generates results, generally described by the term 'explainability', associated with a number of important risks. The widespread use of opaque models could result in unintended consequences, if users of models and supervisors prove unable to predict how the actions directed by ML models could negatively affect the markets. Any intentional lack of transparency by firms in order to protect their advantage adds to the lack of explainability and raises issues related to the supervision of AI algorithms and ML models, but also to the ability of users to adjust their strategies in time of poor performance or in times of stress.

Lack of explainability is incompatible with existing laws and regulations, but also with internal governance, risk management and control frameworks of financial service providers. It limits the

ability of users to understand how their models affect markets or contributes to market shocks, and can amplify systemic risks related to pro-cyclicality. Importantly, the inability of users to adjust their strategies in times of stress may lead to exacerbated market volatility and bouts of illiquidity during periods of acute stress, aggravating flash crash type of events. Explainability issues are deteriorated by a generalised gap in technical literacy and the mismatch between the complexity that is characteristic to AI models and the demands of human-scale reasoning and interpretation that fit the human cognition. Regulatory challenges in terms of transparency and auditing of such models in many financial services use cases.

Financial market practitioners using AI-powered models have to maintain efforts to improve the explainability of such models so as to be able to better comprehend their behaviour in normal market conditions and in times of stress, and manage associated risks. Views differ over the level of explainability that can be reached in AI-driven models, depending on the type of AI used. A fine balance will need to be achieved between interpretability of the model and its level of predictability. The introduction of disclosure requirements around the use of AI-powered models and processes could help mitigate challenges associated with explainability, while also providing more comfort and help build trust in consumers using AI-driven services.

Potential risks should be continually assessed and managed to ensure that AI systems function in a robust and resilient way. The robustness of AI systems can be reinforced by careful training, and retraining, of ML models with datasets large enough to capture non-linear relationships and tail events in the data (including synthetic ones). Ongoing monitoring, testing and validation of AI models throughout their lifecycles, and based on their intended purpose, is indispensable in order to identify and correct for 'model drifts'[2] (concept drifts or data drifts), affecting the model's predictive power. Such model drifts appear when tail events, such as the COVID-19 crisis, give rise to discontinuity in the datasets and are practically difficult to overcome, as they cannot be reflected in the data used to train the model. The role of human judgement remains critical at all stages of AI deployment, from input of datasets to evaluation of model outputs, and can help avoid the risk of interpreting meaningless correlations observed from patterns in activity as causal relationships. Automated control mechanisms or 'kill switches' can also be used as a last line of defence to quickly shut down AI-based systems in case they cease to function according to the intended purpose, although this is also suboptimal as it creates operational risk and assures lack of resilience where the prevailing business system needs to be shut down when the financial system is under stress.

Explicit governance frameworks that designate clear lines of responsibility around AI-based systems throughout their lifecycle, from development to deployment, could further strengthen existing model governance arrangements. Internal model governance committees or model review boards of financial services providers are tasked with the setting of model governance standards and processes for model building, documentation, and validation for any time of model. Such boards are expected to become more common with the wider adoption of AI by financial firms, with possible 'upgrading' of their roles and competencies and some of the processes involved to accommodate for the complexities introduced by AI-based models (e.g. frequency of model validation).

Clear accountability mechanisms are becoming increasingly important, as AI models are deployed in high-value decision-making use-cases (e.g. access to credit). Risks arise also when it comes to outsourcing of AI techniques to third parties, both in terms of accountability and in terms of competitive dynamics (e.g. concentration risk, risk of dependency). Outsourcing of AI models or infrastructure may also give rise to vulnerabilities related to increased risk of convergence related to market positions, which could trigger herding behaviour and convergence in trading strategies and the possibility that large part of the market is affected at the same time, and which could in turn lead to bouts of illiquidity in times of stress.

The technology-neutral approach applied by many jurisdictions to regulate financial market products may be challenged by the rising complexity of some innovative use-cases of AI in finance. Potential inconsistencies with existing legal and regulatory frameworks may arise from the use of advanced

AI techniques (e.g. given the lack of explainability or the adapting nature of deep learning models). Moreover, there may be potential risk of fragmentation of the regulatory landscape with respect to AI at the national, international and sectoral level.

Strengthening of skills sets to develop and manage emerging risks from AI will be needed as AI applications become mainstream in finance. The application of AI by the financial industry may also result in potentially significant job losses across the industry, giving rise to employment challenges.

AI in finance should be seen as a technology that augments human capabilities instead of replacing them. A combination of 'human and machine', where AI informs human judgment rather than replace it (decision-aid instead of decision-maker), could allow for the benefits of the technology to realise, while maintaining safeguards of accountability and control as to the ultimate decision-making. Appropriate emphasis may need to be placed on human primacy in decision making, particularly when it comes to higher-value use-cases (e.g. lending decisions).

Policy considerations

Policy makers and regulators have a role in ensuring that the use of AI in finance is consistent with regulatory aims of promoting financial stability, protecting financial consumers, and promoting market integrity and competition. Policy makers should consider supporting AI innovation in the sector while protecting financial consumers and investors and promoting fair, orderly and transparent markets. Emerging risks from the deployment of AI techniques need to be identified and mitigated to support and promote the use of responsible AI. Existing regulatory and supervisory requirements may need to be clarified and sometimes adjusted, as appropriate, in order to address some of the perceived incompatibilities of existing arrangements with AI applications.

The application of regulatory and supervisory requirements on AI techniques could be looked at under a contextual and proportional framework, depending on the criticality of the application and the potential impact on the consumer outcome and on the market functioning. This will likely encourage the use of AI without unnecessarily stifling innovation. Nonetheless, applying proportionality should not undermine fundamental prudential and stability safeguards, or the protection of investors and financial consumers, all key mandates of policy makers.

Policy makers should consider sharpening their focus on better data governance by financial sector firms, aiming to reinforce consumer protection across AI applications in finance. Specific requirements or best practices for data management in AI-based techniques could be considered, touching upon data quality, adequacy of the dataset used depending on the intended use of the AI model, and safeguards that provide assurance about the robustness of the model when it comes to avoiding potential biases. Appropriate sense checking of model results against baseline datasets and other tests based on whether protected classes can be inferred from other attributes in the data are two examples of best practices to mitigate risks of discrimination. Requirements for additional transparency over the use of personal data and opt-out options for the use of personal data could be considered by authorities.

Policy makers should consider disclosure requirements around the use of AI techniques in the provision of financial services and that it may impact the customer outcome. Financial consumers should be informed about the use of AI techniques in the delivery of a product, as well as potential interaction with an AI system instead of a human being, in order to be able to make conscious choices among competing products. Clear information around the AI system's capabilities and limitations should be included in such disclosure. The introduction of suitability requirements for AI-driven financial services should be considered by authorities to help firms better assess whether prospective clients have a solid understanding of how the use of AI affects the delivery of the product.

Regulators should consider how to overcome the perceived incompatibility of the lack of explainability in AI with existing laws and regulations. There may be a need to update and/or adjust the currently applicable frameworks for model governance and risk management by financial services firms

in order to address such challenges arising by the use of AI-based models. The supervisory focus could be shifted from documentation of the development process and the process by which the model arrives to its prediction to model behaviour and outcomes, and supervisors may wish to look into more technical ways of managing risk, such as adversarial model stress testing or outcome-based metrics (Gensler and Bailey, 2020[2]).

Policy makers should consider requiring clear model governance frameworks and attribution of accountability in order to help build trust in AI-driven systems. Explicit governance frameworks that designate clear lines of responsibility for the development and overseeing of AI-based systems throughout their lifecycle, from development to deployment, could be put in place by financial services providers so as to strengthen existing arrangements for operations related to AI. Internal model governance frameworks could be adjusted to better capture risks emerging from the use of AI, as well as to incorporate intended outcomes for consumers together with an assessment of whether and how such outcomes are reached using AI technologies. Adequate documentation and audit trails of the above processes could assist the oversight of such activity by supervisors.

The provision of increased assurance by financial firms around the robustness and resilience of AI models is fundamental as policy makers seek to guard against build-up of systemic risks, and will help AI applications in finance gain trust. The performance of models needs to be tested in extreme market conditions, to prevent systemic risks and vulnerabilities that may arise in times of stress. The introduction of automatic control mechanisms (such as kill switches) that trigger alerts or switch off models in times of stress could assist in mitigating risks, although they expose the firm to new operational risks. Back-up plans, models and processes should be in place to ensure business continuity in case the models fail or act in unexpected ways. Further, regulators could consider add-on or minimum buffers if banks were to determine risk weights or capital based on AI algorithms (Gensler and Bailey, 2020[2]).

Frameworks for appropriate training, retraining and rigorous testing of AI models could be introduced and/or reinforced to ensure that ML model-based decision-making is operating as intended and in compliance with applicable rules and regulations. Datasets used for training must be large enough to capture non-linear relationships and tail events in the data, even if synthetic, to improve the reliability of such models in times of unpredicted crisis. Continuous testing of ML models is indispensable in order to identify and correct for model drifts.

Regulators should consider promoting the ongoing monitoring and validation of AI models, which are fundamental for their risk, as one of the most effective ways to reinforce model resilience, prevent, and address model drifts. Best practices around standardised procedures for such monitoring and validation could assist in improving model resilience, and identify whether the model necessitates adjustment, redevelopment, or replacement. Model validation, and the necessary approvals and sign-offs should be separated from the development of the model and documented as best possible for supervisory purposes. The frequency of testing and validation would need to be defined, as appropriate, depending on the complexity of the model and the materiality of the decisions made by such model.

Appropriate emphasis could be placed on human primacy in decision making when it comes to higher-value use-cases, such as lending decisions, which significantly affect consumers. Authorities should consider the introduction of processes that can allow customers to challenge the outcome of AI models and seek redress could also help build trust over such systems. The GDPR is an example of such policies, as it provides the associated right of individuals ‘to obtain human intervention’ and to express their points of view if they wish to contest the decision made by an algorithm (EU, 2016[3]).

Policy makers could consider the increased technical complexity of AI, and whether resources will need to be deployed to keep pace with advances in technology. Given the transformative effect of AI on certain financial market activities, as well as the new types of risks stemming from its use, AI has been a growing policy priority for the past few years. Investment should be allocated in research and skills upgrade both for finance sector participants and for enforcement authorities

The role of policy makers is important in supporting innovation in the sector while ensuring that financial consumers and investors are duly protected and the markets around such products and services remain fair, orderly and transparent. Policy makers should consider sharpening their existing arsenal of defences against risks emerging from, or exacerbated by, the use of AI. Clear communication around the adoption of AI and the safeguards in place to protect the system and its users can help instil trust and confidence and promote the adoption of such innovative techniques. Given the ease of cross-border provision of financial services, a multidisciplinary dialogue between policy makers and the industry could be fostered and maintained both at national and international levels.

1 Artificial Intelligence, Machine Learning and Big data in Financial Services

1.1. Introduction

AI systems are machine-based systems with varying levels of autonomy that can, for a given set of human-defined objectives, make predictions, recommendations or decisions using massive amounts of alternative data sources and data analytics referred to as 'big data'[3] (OECD, 2019[4]). Such data feed ML models able to learn from data sets to 'self-improve' without being explicitly programmed by humans.

The COVID-19 crisis has accelerated and intensified the digitalisation trend that was already observed prior to the pandemic, including around the use of AI. Growing AI adoption in finance, in areas such as asset management, algorithmic trading, credit underwriting, and blockchain-based financial services, is enabled by the abundance of available data and by increased, and more affordable, computing capacity.

AI[4] is embedded in products/services across various industries (e.g. healthcare, automobile, consumer products, internet of things (IoT)) and is increasingly being deployed by financial services providers across industries within the financial sector: in retail and corporate banking (tailored products, chat boxes for client service, credit underwriting and scoring, credit loss forecasting, AML, fraud monitoring and detection, customer service); asset management (robo-advice, management of portfolio strategies, risk management); trading (algorithmic trading); insurance (robo-advice, claims management). AI is also being deployed in RegTech and SupTech applications by the official sector (e.g. natural language processing (NLP), compliance processes).

As the deployment of AI and ML using big data is expected to grow in importance (see Section 1.2.1), the possible risks emerging from its application in financial services are becoming more concerning and may warrant further attention by policy makers.

A number of national policy makers and international fora have already launched the debate as to how regulators and supervisors can ensure that the risks stemming from the application of AI in financial services are being mitigated, and what would be the right approach to the deployment of AI in financial services from the policy maker perspective. In other words, how can policy makers support innovation in the sector while ensuring that financial consumers and investors are duly protected and the markets around such products and services remain fair, orderly and transparent?

Given the potentially transformative effect of AI on certain markets, as well as the new types of risks stemming from its use, AI has been a growing policy priority for the past few years. In May 2019, the OECD adopted its Principles on AI (OECD, 2019[5]), the first international standards agreed by governments for the responsible stewardship of trustworthy AI, with guidance from a multi-stakeholder expert group.

The Committee on Financial Markets has included analysis around AI, ML and big data in the Programme of Work and Budget of the Committee for the 2021-22 biennium [C(2008)93/REV2].

This report examines the way AI/ML and big data affect certain financial sector areas that have introduced such technologies early on and how these innovative mechanisms are transforming their business models; discusses benefits and associated risks from the deployment of such technologies in finance; provides an update on regulatory activity and approaches of regulators vis-à-vis AI and ML in financial services in some markets, as well as information on open debates by IOs and other policy makers; identifies areas that

remain concerning and merit further discussion by the Committee and its Experts Group; and provides preliminary policy considerations around these areas. The report does not discuss the use of AI and big data in the insurance sector, which has been discussed by the OECD Insurance and Private Pensions Committee (OECD, 2020[6]).

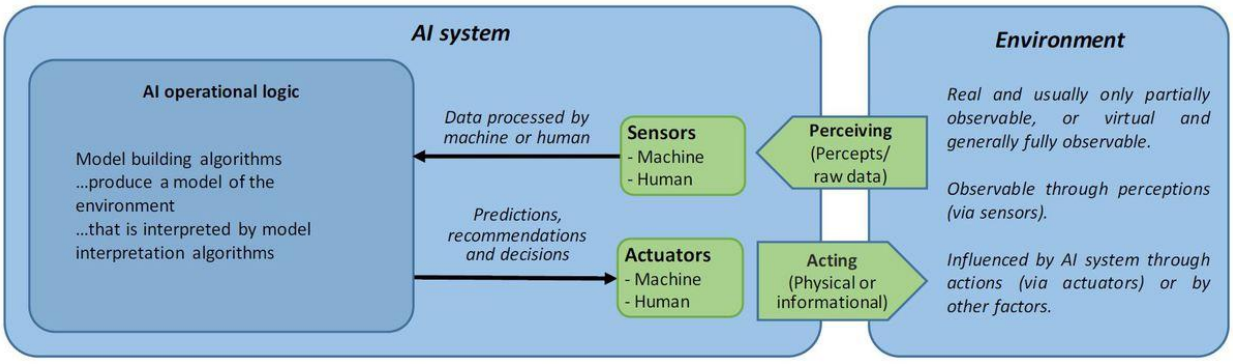
The objective of the discussion and analysis of this topic is twofold: first, to provide analysis to inform the ongoing debate of policy makers and IOs, and second, to explore issues arising in the intersection of AI, finance and policy that remain largely underexplored. The latter involves analysis on how AI, ML and big data influence specific areas of financial market activity (such as asset management; algorithmic trading; credit underwriting; and blockchain-based financial products) and the respective business models; and how such technologies interact with existing risks (such as liquidity, volatility, convergence).

This report has been produced by the Committee's Experts Group on Finance and Digitalisation and has been discussed by the Committee on Financial Markets during the April meetings. Delegates are invited to approve the declassification of this report by written procedure or provide any final comments by **23 July 2021** and approve its publication**.**

1.2. AI systems, ML and the use of big data

An AI system, as explained by the OECD's AI Experts Group (AIGO), is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments (OECD, 2019[4]). It uses machine and/or human-based inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with ML or manually); and use model inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy (OECD, 2019[4]).

Figure 1.1. AI systems



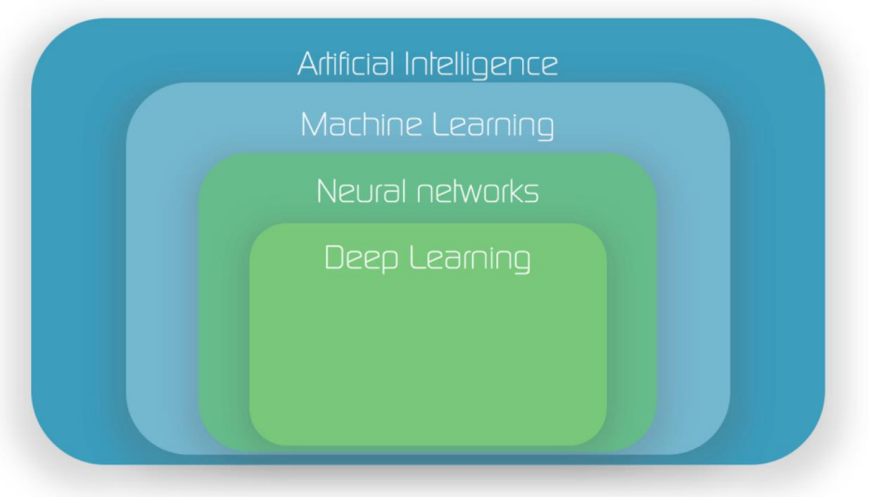
Note: As defined and approved by the OECD AI Experts Group (AIGO) in February 2019.
Source: (OECD, 2019[4]).

The AI system lifecycle phases are (i) planning and design, data collection and processing, and model building and interpretation; (ii) verification and validation; (iii) deployment; and (iv) operation and monitoring (OECD, 2019[4]). An AI research taxonomy distinguishes AI applications (e.g. NLP); techniques to teach AI systems (e.g. neural networks); optimisation (e.g. one-shot-learning); and research addressing societal considerations (e.g. transparency).

ML is an AI subset and describes the ability of software to learn from applicable data sets to 'self-improve' without being explicitly programmed by human programmers (e.g. image-recognition, prediction of borrower default, fraud and AML detection) (Samuel, 1959[7]). The different types of ML include: supervised learning ('classical' ML, consisting of advanced regressions and categorization of data used to improve

predictions) and unsupervised learning (processing input data to understand the distribution of data to develop, for example, automated customer segments); and deep and reinforcement learning (based on neural networks and may be applied to unstructured data like images or voice) (US Treasury, 2018[8]).

Figure 1.2. Illustration of AI subsets



Source: (Hackermoon.com, 2020[9]).

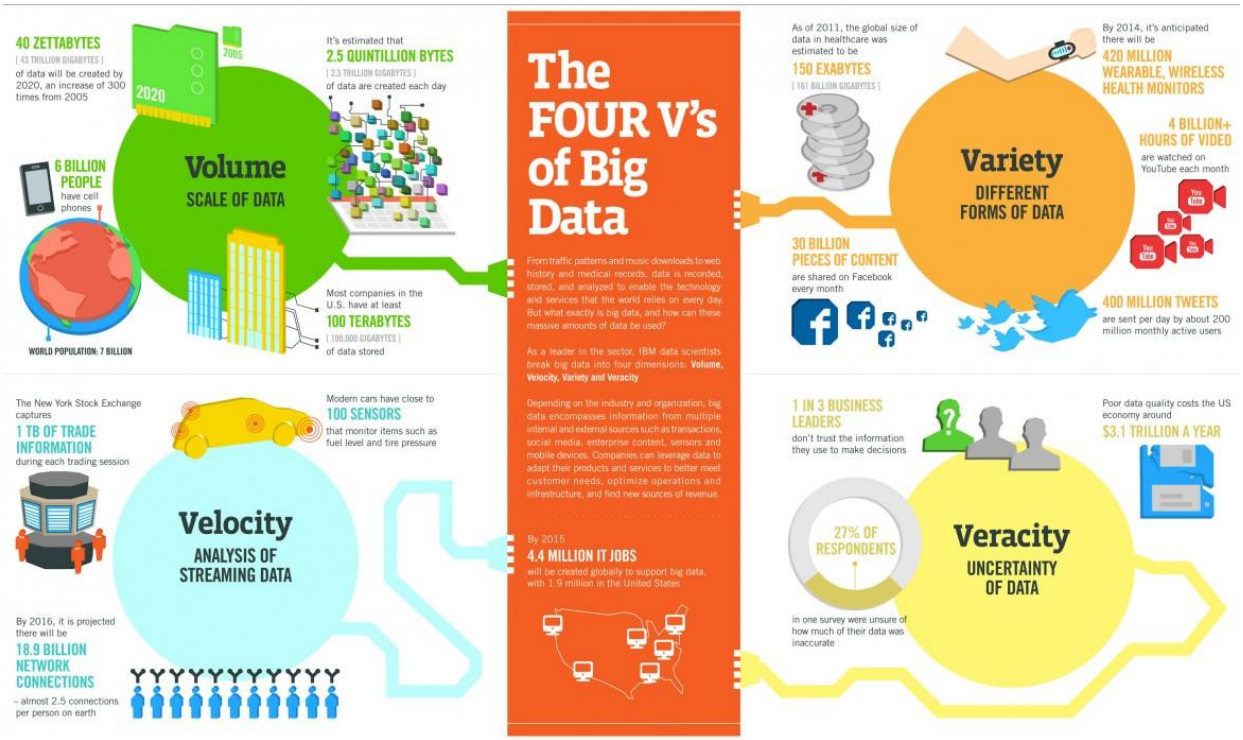
Deep learning neural networks are modelling the way neurons interact in the brain with many ('deep') layers of simulated interconnectedness (OECD, 2019[4]). Such models use multi-layer neural networks[5] to learn and recognise complex patterns in data, inspired by the way the human brain works. Deep learning models can recognise and classify input data without having to write specific rules (no need to specify specific detectors), and can identify new patterns that no human being would have anticipated or developed (Krizhevsky, Sutskever and Hinton, 2017[10]). Such networks are thought to have higher tolerance of noise and can operate at multiple layers of generality from sub features.

ML models use massive amounts of alternative data sources and data analytics that is referred to as 'big data'. The term big data was first coined in the early 2000s when Big Data was used to describe "the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology" (OECD, 2019[4]). The ecosystem of big data encompasses data sources, software, analytics, programming and statistics, and data scientists who synthesise the data to signal out the noise and produce intelligible outputs.

Attributed characteristics of big data include the '4Vs': volume (scale of data); velocity (high-speed processing and analysis of streaming data); variety (heterogeneous data), and veracity (certainty of data, source reliability, truthfulness), as well qualities including exhaustivity, extensionality, and complexity (OECD, 2019[4]) (IBM, 2020[11]). Veracity is of particular importance as it may prove difficult for users to assess whether the dataset used is complete and can be trusted, and may require assessment on a case-by-case basis.

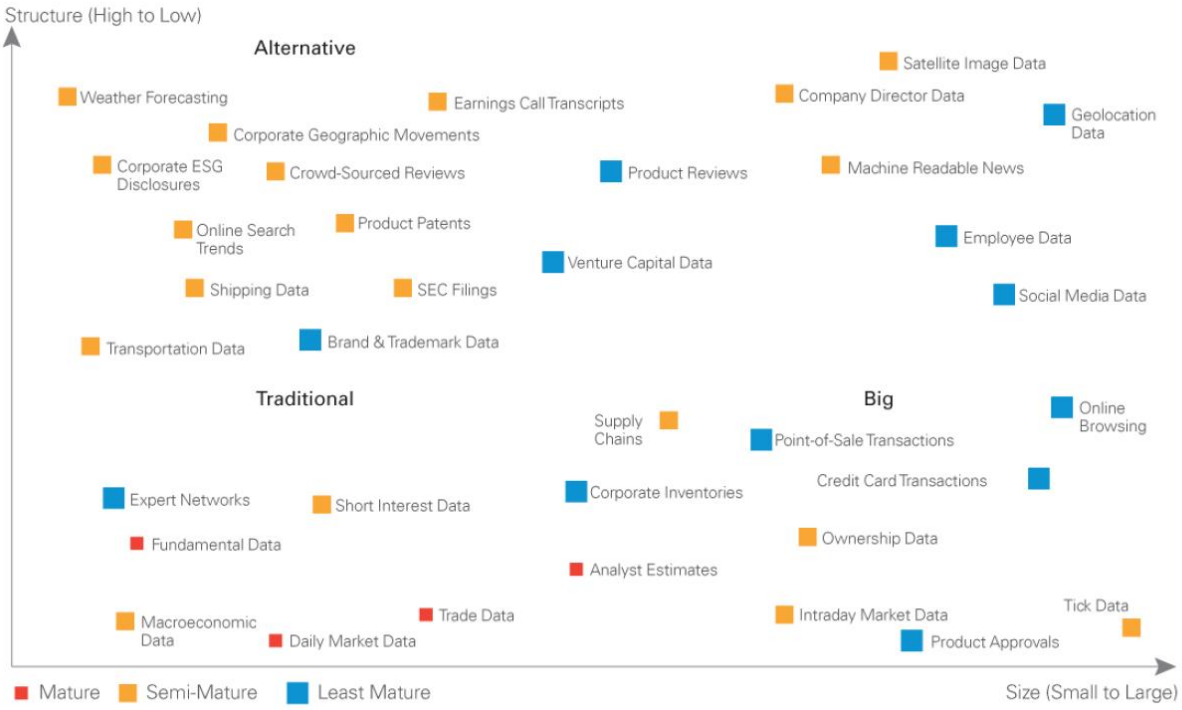
Big data can include climate information, satellite imagery, digital pictures and videos, transition records or GPS signals, and personal data: a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address (OECD, 2019[4]). Such data challenge existing methods due to size, complexity, or rate of availability and requires advanced digital techniques, such as ML models to analyse them. Increased use of AI in IoT applications are also generating significant sums of data, feeding back into AI applications.

Infographic 1.1. The four Vs of Big data



Source: (IBM, 2020[11]).

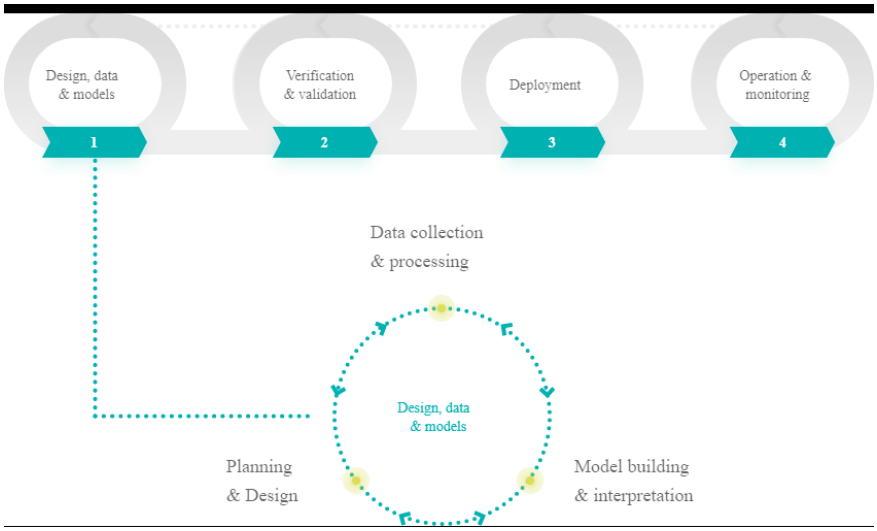
Figure 1.3. Big data sources



Source: Dell Technologies.

Greater data availability allows ML models to perform better because of their ability to learn from the examples fed into the models in an iterative process referred to as training the model (US Treasury, 2018[8]).

Figure 1.4. AI System lifecycle

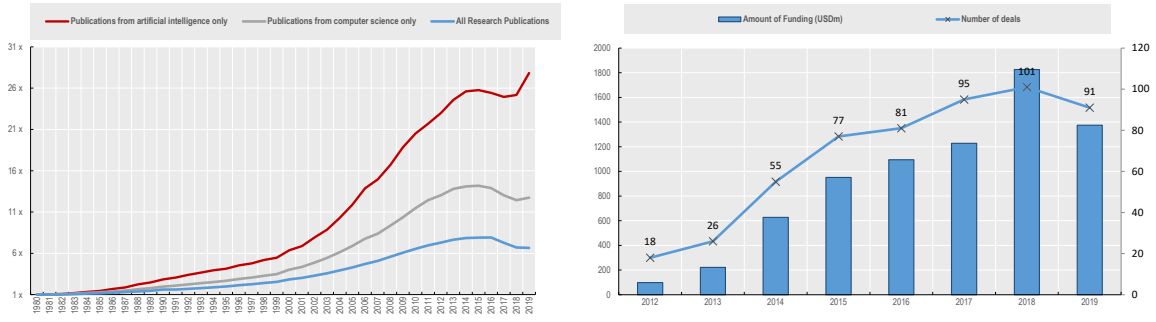


Note: As defined and approved by the OECD AI Experts Group (AIGO) in February 2019.
Source: (OECD, 2019[4]).

1.2.1. A fast-growing area in research and business development

Growth in the deployment of AI applications is evidenced by increased global spending on AI in the private sector, coupled with increased research activity on this technology. Global spending on AI is forecast to double over the next four years, growing from $50.1 billion in 2020 to more than $110 billion in 2024 (OECD, 2019[4]). According to IDC forecasts, spending on AI systems will accelerate over the next several years at an expected CAGR of c.20% for the period 2019-24, as organizations deploy AI as part of their digital transformation efforts and to remain competitive in the digital economy. Private equity investment in AI start-ups doubled in 2017 on a year-to-year basis and attracted 12% of worldwide private equity investments in H1 2018 (OECD, 2019[5]). At the same time, growth in AI-related research is far greater than growth of computer science or overall research publications, providing further evidence of increasing interest around this innovative technology (Figure 2.1).

Figure 1.5. Growth in AI-related research and investment in AI start-ups



Note: Funding of cyber start-ups that use AI as the core product differentiator.
Source: OECD.AI (2020), Microsoft Academic Graph, Insights.

1.2.2. AI in regulatory and supervisory technology ('Regtech' and 'Suptech')

Financial market authorities are increasingly looking into potential benefits from the use of AI insights in 'Suptech' tools, i.e. in FinTech-based applications used by authorities for regulatory, supervisory and oversight purposes (FSB, 2020[12]). Equally, regulated institutions are developing and adopting FinTech applications for regulatory and compliance requirements and reporting ('RegTech'). Financial institutions are adopting AI applications for internal controls and risk management, too, and combination of AI technologies with behavioural sciences allows large financial institutions to prevent misconduct, shifting the focus from ex-post resolution to forward-looking prevention (Scott, 2020[13]).

The growth in RegTech and SupTech applications is mainly attributed to both supply side drivers (increased availability of data, including machine-readable ones, development of AI techniques) and demand side drivers (potential for gains in efficiency and effectiveness of regulatory processes, possibility for improved insights into risk and compliance developments) (FSB, 2020[12]).

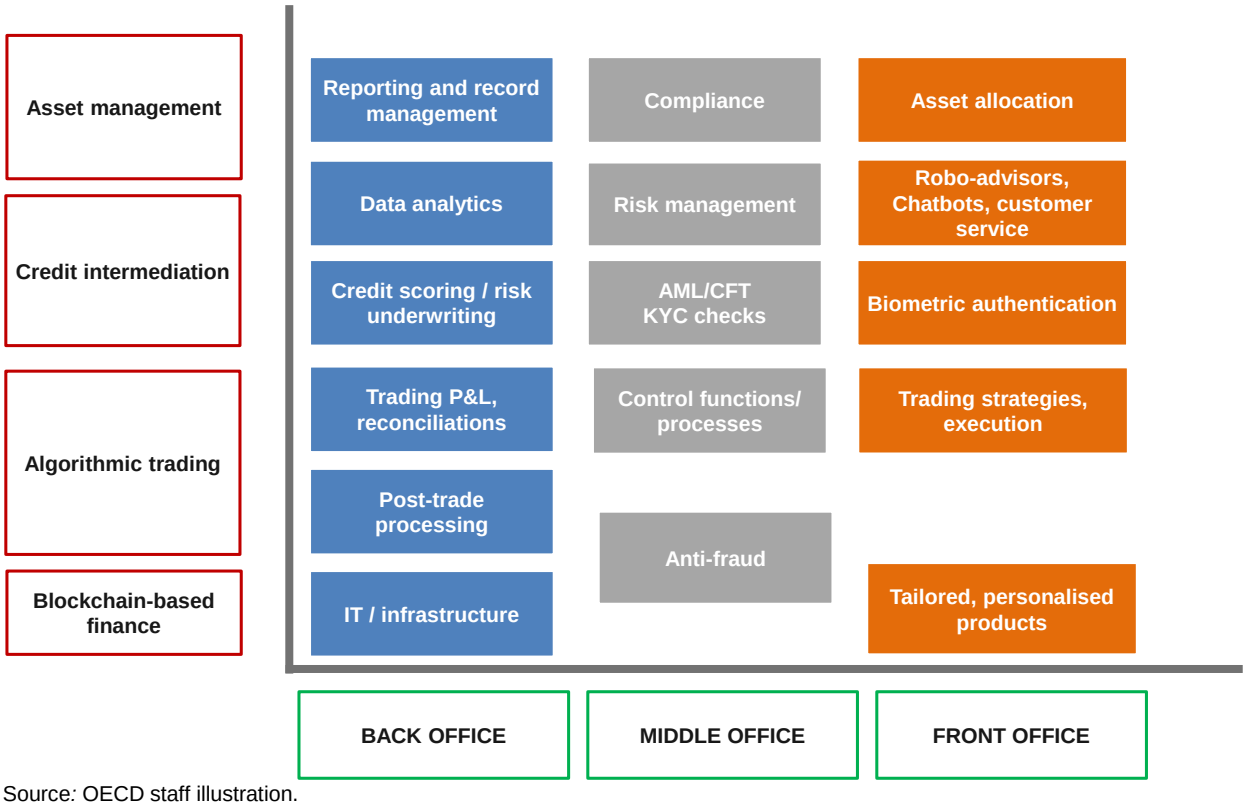
Despite the opportunities and benefits of the application of AI for regulatory and supervisory purposes, authorities remain vigilant given risks associated to the use of such technologies (resourcing, cyber risk, reputational risk, data quality issues, limited transparency and interpretability) (FSB, 2020[12]). These are also the risks prevailing in the deployment of AI by financial market participants, and which are discussed in more detail in this report.

2 AI/ML, big data in finance: benefits and impact on business models/activity of financial sector participants

The adoption of AI in finance is driven by the large and growing availability of data within financial services and the expected competitive advantage that AI/ML can provide to financial services firms. The explosion in the abundance of available data and analytics (big data), coupled with more affordable computing capacity (e.g. cloud computing) can by analysed by ML models to identify signals and capture underlying relationships in data in a way that is beyond the ability of humans. The deployment of AI/ML and big data by financial sector companies is expected to increasingly drive firms' competitive advantage, through both improving the firms' efficiency by reducing costs, and enhancing the quality of financial services products demanded by customers (US Treasury, 2020).

This section looks at the potential impact that the use of AI and big data may have in specific financial market activities, including asset management and investing, trading; lending; and blockchain applications in finance.

Figure 2.1. Examples of AI applications in some financial market activities



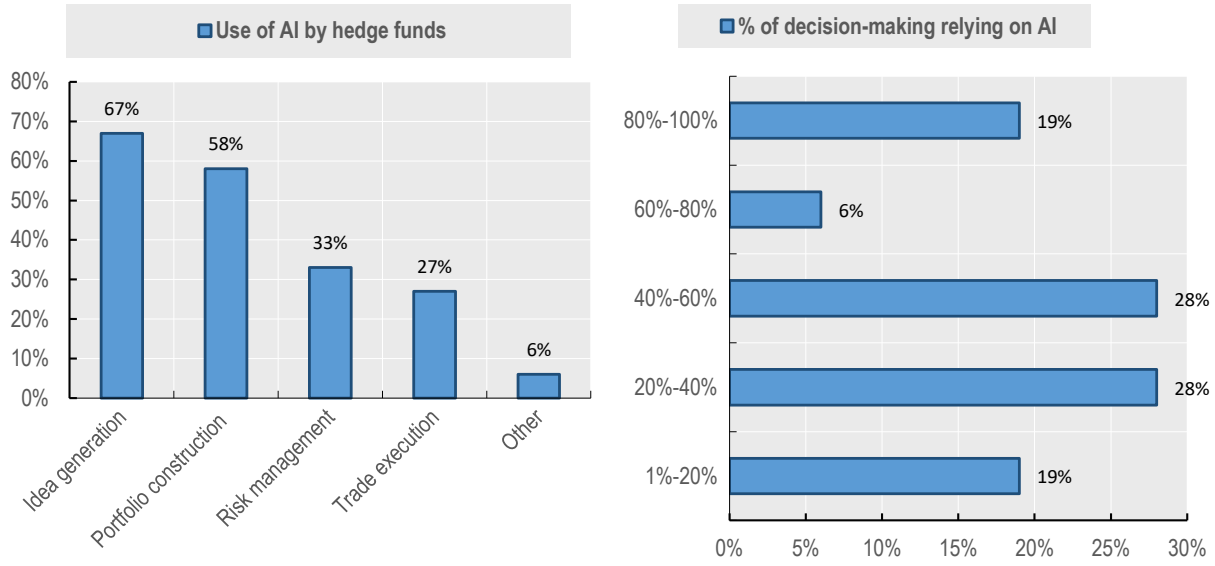
Source: OECD staff illustration.

2.1. Portfolio allocation in asset management[6] and the broader investment community (buy-side)

The use of AI and ML in asset management has the potential to increase the efficiency and accuracy of operational workflows, enhance performance, strengthen risk management, and improve the customer experience (Blackrock, 2019[14]) (Deloitte, 2019[15]). Natural Language Generation (NLG), a subset of AI, can be used by financial advisors to 'humanise' and simplify data analysis and reporting to clients (Gould, 2016[16]). As ML models can monitor thousands of risk factors on a daily basis and test portfolio performance under thousands of market/economic scenarios, the technology can enhance risk management for asset managers and other large institutional investors. In terms of operational benefits, the use of AI can reduce back-office costs of investment managers, replace manually intensive reconciliations with automated ones, and potentially reduce costs and increase speed.

Feeding ML models with big data can provide asset managers with recommendations that influence decision-making around portfolio allocation and/or stock selection, depending on the type of AI technique used. Big data has replaced traditional datasets, which are now considered a commodity easily available to all investors, and is being used by asset managers to gain insights in their investment process. For the investment community, information has always been key and data has been the cornerstone of many investment strategies, from fundamental analysis to systematic trading and quantitative strategies alike. While structured data was at the core of such 'traditional' strategies, vast amounts of raw or unstructured/semi-structured data are now promising to provide a new informational edge to investors deploying AI in the implementation of their strategies. AI allows asset managers to digest vast amounts of data from multiple sources and unlock insights from the data to inform their strategies at very short timeframes.

Figure 2.2. AI use by hedge funds (H1 2018)



Note: Based on industrial research by Barclays, as of July 2018.
Source: (BarclayHedge, 2018[17]).

The use of AI/ML and big data may be reserved to larger asset managers or institutional investors who have the capacity and resources to invest in AI technologies, possibly introducing a barrier for the adoption of such techniques by smaller actors. Investment in technology and in talent is required to transform and explore vast amounts of unstructured new datasets of big data and build ML models. To the extent that the deployment of AI and proprietary models provides a performance edge against competition, this may, in turn, result in restricted participation by smaller players who cannot adopt in-house AI/ML techniques or use big data information sources. This could potentially reinforce the trend of concentration in a small number of larger players that is being observed in the hedge fund industry, as bigger groups outpace some of their more nimble rivals (Financial Times, 2020[18]).

Restricted participation by smaller players would persevere at least until the industry reaches a point where such tools become ubiquitous/provided as a service by third party vendors. At the same time, third party datasets may not be held at the same standard across the industry, and users of third party tools will have to build confidence as to the accuracy and trustworthiness of data used ('veracity' of big data) so as to reach a level of comfort sufficient for them to adopt them.[7]

The use of the same AI models by a large number of asset managers could lead to herding behaviour and one-way markets, which may raise potential risks for liquidity and the stability of the system particularly in times of stress. Market volatility could increase through large sales or purchases executed simultaneously, giving rise to new sources of vulnerabilities (see Section 2.2).

It could be argued that the deployment of AI/ML and big data in investing could somehow reverse the trend towards passive investing. If the use of such innovative technologies proves to be alpha generating in a consistent manner that suggests some level of cause-and-effect relationship between the use of AI and the superior performance (Blackrock, 2019[14]) (Deloitte, 2019[15]), the active investment community could leverage this opportunity to reinvigorate active investing and provide alpha-adding opportunities to their clients.

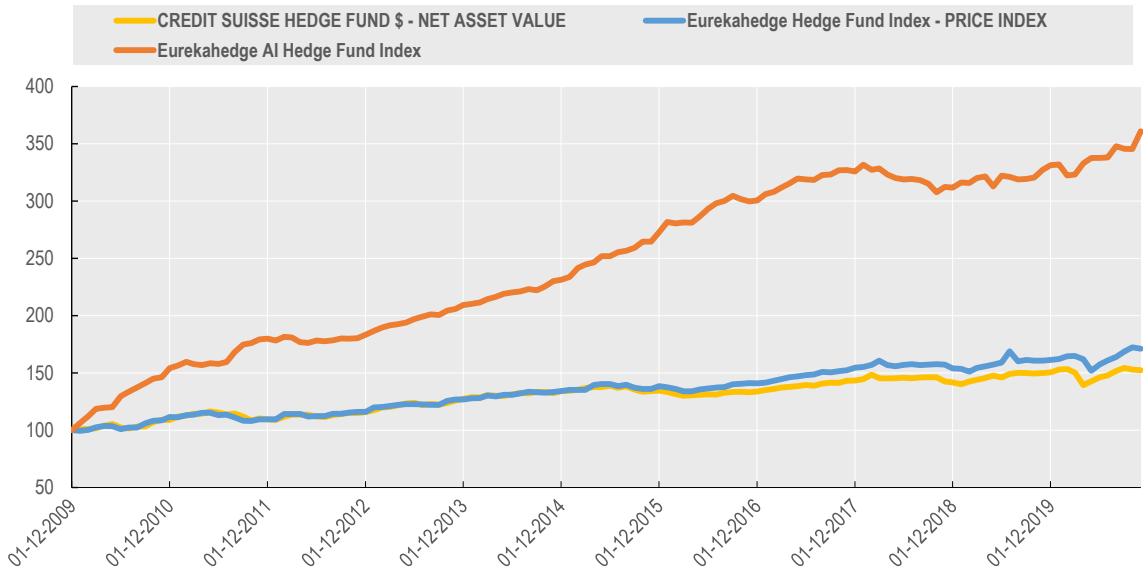
2.1.1. Performance of AI-powered hedge funds and ETFs

Hedge funds have been on the leading edge of FinTech users, and use big data, AI and ML algorithms in trade execution and back office functions (Kaal, 2019[19]). A class of 'AI pure play' hedge funds has emerged in recent years that are based entirely on AI and ML (e.g. Aidiyia Holdings, Cerebellum Capital, Taaffeite Capital Management and Numerai) (BNY Mellon, 2019[20]).

To date, there has been no academic or other independent review of the performance of AI-powered funds from a non-industry source, comparing the various funds claiming to be AI-driven. Fund managers use different degrees of AI in their operations and strategies and are naturally reluctant to disclose their methodologies so as to maintain their competitive advantage. While many funds may be marketing their products as 'AI powered', the levels at which such technologies are used by funds and the maturity of deployment of AI vary significantly, therefore making it difficult to compare performance between the different self-proclaimed AI products.

Private sector provided indices of AI-powered hedge funds demonstrate outperformance of AI-based funds over conventional hedge fund indices provided by the same source (Figure 2.2). Indices provided by third parties are prone to a number of biases, such as survivorship and self-selection bias of constituents to the index or back filling, and should be treated with caution.

Figure 2.3.Some AI-powered hedge funds have outperformed conventional hedge funds



Note: The *Eurekahedge Hedge Fund Index* is Eurekahedge's flagship equally weighted index of 2195 constituent funds. The index is designed to provide a broad measure of the performance all underlying hedge fund managers irrespective of regional mandate. The index is base weighted at 100 at December 1999, does not contain duplicate funds and is denominated in local currencies. The *Eurekahedge AI Hedge Fund Index* is an equally weighted index of 18 constituent funds. The index is designed to provide a broad measure of the performance of underlying hedge fund managers who utilize AI and ML theory in their trading processes. The index is base weighted at 100 at December 2010, does not contain duplicate funds and is denominated in USD. The *Credit Suisse Hedge Fund Index* is an asset-weighted hedge fund index and includes open and closed funds.
Source: Eurekahedge; Datastream, Thompson Reuters Eikon.

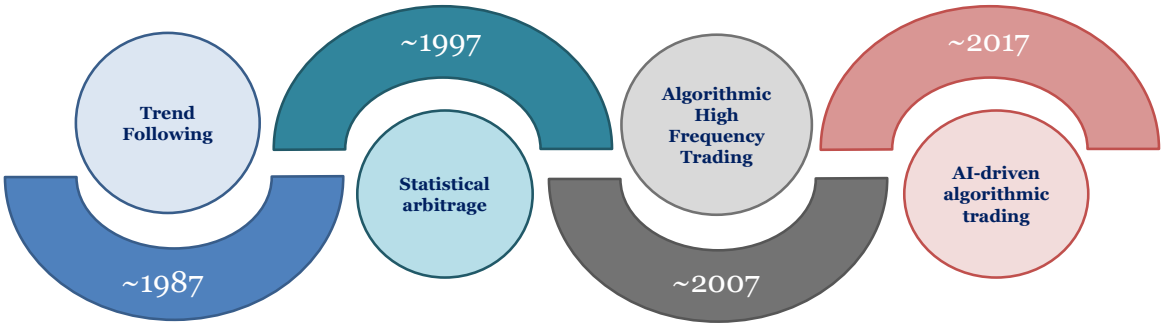
ETFs powered by AI, in which investment decisions are made and executed by models, have not reached a meaningful size as of yet. The total AuM of this cohort of ETFs has been estimated to stand at c. USD 100 m as of end of 2019 (CFA, n.d.[21]). The efficiencies produced by the deployment of AI on automated ETFs lowers management fees (estimated at an average yearly fee of 0.77% as of end of 2019). In terms of forecasting performance, there is increasing evidence that ML models outperform conventional forecasts of macroeconomic indicators such as inflation and GDP (Kalamara et al., 2020[22]). In fact, these improvements are most pronounced during periods of economic stress when, arguably, forecasts matter most. Evidence also exists on the superiority of AI-driven techniques in identifying meaningful but previously unknown correlations in the pattern of financial crises, with ML models mostly outperforming logistic regression in out-of-sample predictions and forecasting (Bluwstein et al., 2020[23]).

2.2. Algorithmic Trading

AI can be used in trading both to provide trading strategy suggestions and to power automated trading systems that make predictions, choose the course of action and execute trades. AI-based trading systems operating in the market can identify and execute trades entirely on their own, without any human intervention, using AI techniques such as evolutionary computation, deep learning and probabilistic logic (Metz, 2016[24]). AI techniques (such as algo wheels[8]) can help strategize any upcoming trade in a systematic fashion by enabling an "if/then" thought process to be implemented as a matter of procedure (Bloomberg, 2019[25]) (see Box 2.1). Given today's interconnectedness between asset classes and geographies, the use of AI allows for predictive capacity that is fast outpacing the power of even conventional algos in finance and trading.

AI-enabled systems in trading can also assist traders in their risk management and in the management of the flow of their orders. For example, AI-based applications can track the risk exposure and adjust or exit the position depending on the needs of the user, in a fully automated manner and without the need for reprogramming, as they train on their own and adapt to changing market circumstances without (or with minimal) human intervention. They can help traders manage their flows among their brokers, for trades that are already decided upon, and control fees or liquidity allocation to different pockets (e.g. regional market-preferences, currency determinations or other parameters of an order handling) (Bloomberg, 2019[25]).

Figure 2.4. Historical evolution of trading and AI



In highly digitised markets, such as the ones for equities and FX products, AI solutions promise to provide competitive pricing, manage liquidity, optimise and streamline execution. Importantly, AI algorithms deployed in trading can enhance liquidity management and execution of large orders with minimum market impact, by optimising size, duration and order size in a dynamic fashion, based on market conditions.

The use of AI and big data in sentiment analysis to identify themes, trends, and trading signals is augmenting a practice that is not new. Traders have mined news reports and management announcements/commentaries for decades now, seeking to understand the stock price impact of non-financial information. Today, text mining and analysis of social media posts and twits or satellite data through the use of NPL algorithms is an example of the application of innovative technologies that can inform trading decisions, as they have the capacity to automate data gathering and analysis and identify persistent patterns or behaviours on a scale that a human cannot process.

What differentiates AI-managed trading with systematic trading is the reinforcement learning and adjustment of the AI model to changing market conditions, when traditional systematic strategies would take longer to adjust parameters due to the heavy human intervention involved. Conventional back-testing strategies based on historical data may fail to deliver good returns in real time as previously identified trends break down. The use of ML models shifts the analysis towards prediction and analysis of trends in real time, for example using 'walk forward' tests[9] instead of back testing.[10] Such tests predict and adapt to trends in real time to reduce over-fitting (or curve fitting, see Section 3.5.1.) in back tests based on historical data and trends (Liew, 2020[26]).

Box 2.1. AI-based algo wheels

An algo wheel is a broad term, encompassing fully automated solutions to mostly trader-directed flow. An AI-based algo wheel is an automated routing process embedding AI techniques to assign a broker algorithm to orders from a pre-configured list of algorithmic solutions (Barclays Investment Bank, 2020[27]). In other words, AI-based algo wheels are models that select the optimal strategy and broker through which to route the order, depending on market conditions and trading objectives/requirements.

Investment firms typically use algo wheels for two reasons; first, to achieve performance gains from improved execution quality; second, to gain workflow efficiency from automating small order flow or normalizing broker algorithms into standardized naming conventions. Market participants argue that algo wheels reduce the trader bias around the selection of the broker and broker's algorithm deployed in the marketplace.

It is estimated that almost 20% of trading flows are currently going through algo wheels, and the mechanism is increasingly gaining acceptance as a way of systematically categorizing and measuring the best performing broker algos (Mollemans, 2020[28]). However, those who do, use it for 38% of their algo flow. A potential wide adoption of algo wheels could therefore lead to an increase in the overall level of electronic trading, with potential benefits for the competitive landscape of electronic brokerage (Weber, 2019[29]).

The use of AI in trading has gone through different stages of development and corresponding complexity, adding a layer to traditional algorithmic trading at each step of the process. First-generation algorithms consisted of buy or sell orders with simple parameters, followed by algorithms allowing for dynamic pricing. Second-generation algorithms deployed strategies to break up large orders and reduce potential market impact, helping obtain better prices (so-called 'execution algos'). Current strategies based on deep neural networks are designed to provide the best order placement and execution style that can minimize market impact (JPMorgan, 2019[30]). Deep neural networks mimic the human brain through a set of algorithms designed to recognise patterns, and are less dependent on human intervention to function and learn (IBM, 2020[31]). The use of such techniques can be beneficial for market makers in enhancing the management of their inventory and reduce the cost of their balance sheet. As the development of AI advances, AI algorithms evolve into automated, computer programmed algorithms that learn from the data input used and rely less on human intervention.

In practice, the more advanced forms of AI today are mostly used to identify signals from 'low informational value' incidents in flow-based trading[11], which consist of less obvious events, harder to identify and extract value from. Rather than help with speed of execution, AI is actually used to extract signal from noise in data and converts this information into decision about trades. Less advanced algorithms are mostly used for 'high informational events', which consist of news of financial events that are more obvious for all participants to pick up and where execution speed is of the essence.

At this stage of their development, ML-based models are therefore not aiming at front-running trades and profit from speed of action, such as HFT strategies. Instead, they are mostly confined to being used offline, for example for the calibration of algorithm parameters and for improving algorithms' decision logic, rather than for execution purposes (BIS Markets Committee, 2020[32]). In the future, however, as AI technology advances and is deployed in more use cases, it could amplify the capabilities of traditional algorithmic trading, with implications for financial markets. This is expected to occur when AI techniques start getting deployed more at the execution phase of trades, offering increased capabilities for automated execution of trades and serving the entire chain of action from picking up signal, to devising strategies, and executing them. ML-based algos for execution will allow for the autonomous and dynamic adjustment of their own decision logic while trading. In that case, the requirements already applied for algorithmic trading (e.g.

safeguards built in pre-trading risk management systems, automated control mechanisms to switch off the algorithm when it goes beyond the limits embedded in the risk model) should be extended to AI-driven algorithmic trading.

2.2.1. Unintended consequences and possible risks

The use of the same or similar models by a large number of traders could have unintended consequences for competition, and could also contribute to the amplification of stress in markets. For traders, the emergence of widely-used models would naturally reduce the arbitrage opportunities available, driving down margins. This would ultimately benefit consumers by reducing bid-ask spreads. At the same time, it could potentially result in convergence, herding behaviour and one-way markets, with possible implications for the stability of the market and for liquidity conditions particularly during periods of acute stress. As with any algorithm, wide use of similar AI algorithms creates the risk of self-reinforcing feedback loops that can, in turn, trigger sharp price moves (BIS Markets Committee, 2020[32]).

Such convergence could also increase the risk of cyber-attacks, as it becomes easier for cyber-criminals to influence agents acting in the same way rather than autonomous agents with distinct behaviour (ACPR, 2018[33]). When it comes to cyber risk, when AI is used in a nefarious manner, it has the potential to offensively conduct autonomous attacks (without human intervention) on vulnerable systems in trading but also broadly in financial market systems and participants (Ching TM, 2020[34]).

The use of proprietary models that cannot be copied is key for traders to retain any advantage, and may drive intentional lack of transparency, adding to the lack of explainability of ML models. Such unwillingness by users of ML techniques to reveal their model workings for fear of losing their competitive edge raises also issues related to the supervision of algorithms and ML models (see Section 3.4).

The use of algorithms in trading can also make collusive outcomes easier to sustain and more likely to be observed in digital markets (OECD, 2017[35]) (see Section 4.2.1). Related to that is the risk that AI-driven systems may exacerbate illegal practices aiming to manipulate the markets, such as ‘spoofing’, by making it more difficult for supervisors to identify such practices if collusion among machines is in place (see Box 3.2. The lack of explainability of ML models used to back trading could make the adjustment of the strategy difficult in times of poor trading performance. Trading algorithms are no longer model-based linear processes (input A caused trading strategy B to be executed) that can be traced and interpreted and where there can be a clear understanding of which parameters drove the outcomes. In times of poor performance, it is crucial for traders to be able to decompose the output into the underlying drivers of the trading decision, so as to adjust and/or correct according to the circumstances. However, even in times of over performance, users are unable to understand why the successful trading decision was made, and therefore cannot identify whether the performance is due to the model’s superiority and ability to capture underlying relationships in the data or to pure luck.

In terms of potential unintended effects in the market, it could be argued that the application of AI technologies in trading and HFT could increase market volatility through large sales or purchases executed simultaneously, giving rise to new sources of vulnerabilities (Financial Stability Board, 2017[36]). In particular, some algo-HFT strategies appear to have contributed to extreme market volatility, reduced liquidity and exacerbated flash crashes that have occurred with growing frequency over the past several years (OECD, 2019[37]). As HFT are a major source of liquidity provision under normal market conditions, improving market efficiency, any disruption in the operation of their models in times of crisis results in liquidity being pulled out of the market, with potential impact on market resilience.

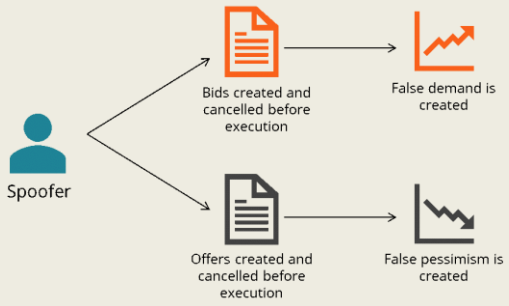
Box 2.2. "Spoofing": the use of algorithms for market manipulation

Spoofing is an illegal market manipulation practice that involves placing bids to buy or offers to sell securities or commodities with the intent of cancelling the bids or offers prior to the deal's execution. It is designed to create a false sense of investor demand in the market, thereby manipulating the behaviour and actions of other market participants and allowing the spoofer to profit from these changes by reacting to the fluctuations.

Spoofing has been possible in trading before the advent of algorithmic trading, but became prominent with the rise of high frequency trading. Market manipulation using spoofing schemes was determined as one of the primary triggers of the 2010 Flash Crash (US Department of Justice, 2015[38]).

In a hypothetical scenario, deep learning ML models that learn from the behaviour of other models and adapt to the circumstances could begin to collude with other ML models to take advantage of such practices. In such cases, a trading entity using ML models may become involved in spoofing and rather than benefit for itself, may implicitly pass on the benefit to another model of the firm or even another trading entity using similar models, potentially making it more difficult for supervisors to identify and prove intent. This can be achieved as ML models can coordinate parallel behaviour without actually engaging in explicit communication, and self-learning and reinforcement learning models learn and dynamically re-adapt their behaviour to the actions of other players.

Figure 2.5. Spoofing practices



Source: CFI.

Similar to the considerations discussed in investing, the possible massive use of 'off-the-shelf' AI models by market participants could have potential effects for liquidity and market stability, by prompting herding and one-way markets. Such behaviour would also amplify volatility risks, pro-cyclicality and unexpected changes in the market both in terms of scale and in terms of direction. Herding behaviour may lead to illiquid markets in the absence of 'shock-absorbers' or market makers available and able to take on the opposite side of transactions.

The deployment of AI in trading may also increase the interconnectedness of financial markets and institutions in unexpected ways, and potentially increase correlations and dependencies of previously unrelated variables (Financial Stability Board, 2017[36]). The scaling up of the use of algorithms that generate uncorrelated profits or returns may generate correlation in unrelated variables if their use reaches a sufficiently important scale. It can also amplify network effects, such as unexpected changes in the scale and direction of market moves.

In order to mitigate risks from the deployment of AI in trading, defences may need to be put in place for AI-driven algorithmic trading. Safeguards built in pre-trading risk management systems aim to prevent and stop potential misuse of such systems. Interestingly, AI is also being used to build better pre-trade risk

systems, which include, inter alia, mandatory testing of every release of an algo, and which would apply equally to AI-based algorithms. Automated control mechanisms that instantly switch off the model are the ultimate lines of defence of market practitioners, when the algorithm goes beyond the risk system, and consist of 'pulling the plug' and replacing any technology with human handling.[12] Such mechanisms could be considered suboptimal from a policy perspective, as they switch off the operation of the systems when it is most needed in times of stress, and give rise to operational vulnerabilities.

Defences may also need to be applied at the level of the exchanges where the trading is taking place. These could include automatic cancellation of orders when the AI system is switched off for some reason and methods that provide resistance to sophisticated manipulation methods enabled by technology. Circuit breakers, currently triggered by massive drops between trades, could perhaps be adjusted to also identify and be triggered by large numbers of smaller trades performed by AI-driven systems, with the same effect.

2.3. Credit intermediation and assessment of creditworthiness

AI-based models and big data are increasingly being used by banks and fintech lenders to assess the creditworthiness of prospective borrowers and make underwriting decisions, both functions at the core of finance. In the context of credit scoring, ML models are used to predict borrowers' defaults with superior forecasting accuracy compared to standard statistical models (e.g. logic regressions) especially when limited information is available (Bank of Italy, 2019[39]) (Albanesi and Vamossy, 2019[40]). Moreover, financial intermediaries use AI-based systems for fraud detection, as well as in order to analyse the degree of interconnectedness between borrowers, which in turn allows them to better manage their lending portfolio.

Box 2.3. AI for fraud detection, transaction screening and monitoring

AI and big data are being used in fraud detection by financial institutions and FinTech lenders, for client on boarding and KYC checks, anti-money laundering and terrorist financing screening on a shared platform at on boarding and ongoing customer due diligence stage (AML/CFT) and detection of suspicious activities during ongoing monitoring.

In particular, AI can help institutions recognise abnormal transactions and identify suspicious and potentially fraudulent activity through the use of image recognition software, risk models, and other AI-based techniques (e.g. fraudulent use of customer's personal information, misrepresenting products/services, other scams). AI can also reduce the incidence of false positives, in other words the rejection of otherwise valid transactions (e.g. credit card payment falsely declined), resulting in higher client satisfaction.

A Proof of Concept project to assess the feasibility and effectiveness of the use of AI in AML/CFT on a shared platform has been conducted in Japan. The AI based system for transaction screening and monitoring, using previously filed suspicious transactions from various financial institutions as ML data as objective function, successfully assisted compliance personals in triaging the results of transaction screening against sanctions lists and identifying suspicious transactions (New Energy and Industrial Technology Development Organization, 2021[41]).

At the same time, other applications of AI could be used to circumvent fraud detection capabilities of financial institutions. For example, AI-based fraudulent pictures can become indistinguishable from actual pictures, raising important challenges to authentication and verification functions within financial services (US Treasury, 2018[8]).

The availability of big data and advanced AI-based analytics models using such datasets have transformed the way credit risk is assessed. Credit scoring models powered by AI combine the use of conventional credit information, where available, with big data not intuitively related to creditworthiness (e.g. social media data, digital footprints, and transactional data accessible through Open Banking initiatives).

The use of AI models in credit scoring can reduce the cost of underwriting, while allowing for the analysis of creditworthiness of clients with limited credit history ('thin files'). It can therefore enable the extension of credit to viable companies that cannot prove their viability through historical performance data or tangible collateral assets, potentially enhancing access to credit and supporting the growth of the real economy by alleviating constraints to SME financing. Recent empirical analysis could even reduce the need for collateral by reducing information asymmetries prevailing in credit markets (BIS, 2020[42]). Credit approval rates for parts of the population that has historically been left behind, such as near-prime clients or underbanked parts of the population, could be better served through alternative scoring methods, potentially promoting financial inclusion. Notwithstanding the above, AI-based credit scoring models remain untested over longer credit cycles or in case of a market downturn, and there is limited conclusive empirical support as to the benefits of ML-driven techniques for financial inclusion. For example, while some analysis suggests that the use of ML models for credit risk assessment results in cheaper access to credit only for majority ethnic groups (Fuster et al., 2017[43]), others find that lending-decision rules based on ML predictions help reduce racial bias in the consumer loan market (Dobbie et al., 2018[44]).

2.3.1. AI/ML-based credit scoring, transparency and fairness in lending

Despite their vast potential for speed, efficiency and risk scoring of the 'unscored', AI/ML-based models raise risks of disparate impact in credit outcomes and the potential for discriminatory or unfair lending (US Treasury, 2016[45]).[13] Similar to other applications of AI in finance, such models also raise important challenges related to the quality of data used and the lack of transparency/explainability around the model.

Well-intentioned ML models may inadvertently generate biased conclusions, discriminated against certain classes of people (e.g. based on race, gender, ethnicity, religion) (White & Case, 2017[46]). Inadequately designed and controlled AI/ML models carry a risk of exacerbating or reinforcing existing biases while making discrimination in credit allocation even harder to find (Brookings, 2020[47]).

As with any model used in financial services, the risk of 'garbage in, garbage out' exists in AI/ML-based models for risk scoring and beyond. Inadequate data may include poorly labelled or inaccurate data, data that reflects underlying human prejudices, or incomplete data (S&P, 2019[48]). A neutral ML model that is trained with inadequate data, risks producing inaccurate results even when fed with 'good' data. Alternately, a neural network trained on high-quality data, which is then fed inadequate data, will produce a questionable output, despite the well-trained underlying algorithm. This, combined with the lack of explainability in ML models, makes it harder to detect inappropriate use of data or use of unsuitable data in AI-based applications.

As such, the use of poor quality or inadequate/unsuitable data may result in wrong or biased decision-making. Biased or discriminatory scoring may not be intentional from the perspective of the firm using the model; instead, algorithms may combine facially neutral data points and treat them as proxies for immutable characteristics such as race or gender, thereby circumventing existing non-discrimination laws (Hurley, 2017[49]). For example, while a credit officer may be diligent not to include gender-based variants as input to the model, the model can infer the gender based on transaction activity, and use such knowledge in the assessment of creditworthiness, circumventing the law. Biases may also be inherent in the data used as variables and, given that the model trains itself on data from external sources that may have already incorporated certain biases, perpetuates historical biases.

Similar to other applications of AI in finance, ML-based models raise issues of transparency given their lack of explainability, i.e., the difficulty in comprehending, following or replicating the decision-making process (see Section 3.4). Issues related to explainability are particularly pertinent in lending decisions, as

lenders are accountable for their decisions and must be able to explain the basis for denials of credit extension. This also means that consumers have limited ability to identify and contest unfair credit decisions, and little chance to understand what steps they should take to improve their credit rating.

Regulations in developed economies ensure that specific data points are not taken into account in the credit risk analysis (e.g. US regulation around race data or zip code data, UK regulation around protected category data). Regulation promoting anti-discrimination principles, such as the US fair lending laws, exists in many jurisdictions, and regulators are globally considering the risk of potential bias and discrimination risk that AI/ML and algorithms can pose (White & Case, 2017[46]).

In some jurisdictions, comparative evidence of disparate treatment, such as lower average credit limits for members of protected groups than for members of other groups, is considered discrimination regardless of whether there was intent to discriminate. Potential mitigants against such risks are the existence of auditing mechanisms that sense check the results of the model against baseline datasets; testing of such scoring systems to ensure their fairness and accuracy (Citron and Pasquale, 2014[50]); and governance frameworks for AI-enabled products and services and assignment of accountability to the human parameter of the project, to name a few.

2.3.2. BigTech and financial services

As BigTech increasingly leverage their free access to vast amounts of customer data that feed into AI-driven models to provide financial services, their deployment of AI raises issues around data privacy and concerns over ways in which the collection, storage and use of personal data may be exploited for commercial gain (DAF/CMF(2019)29/REV1). These practices could disadvantage customers, such as through discriminatory practices related to credit availability and pricing.

Access to customer data by BigTech give them a clear competitive advantage over conventional financial services providers. This advantage is likely to be further reinforced with their use of AI, which offers possibilities for novel, customised and more efficient service provision by these players. The dominance of BigTech in certain areas of the market could lead to excessive market concentration (see also Section 3.2) and increase the dependence of the market to few large BigTech players, with possible systemic implications depending on their scale and scope (DAF/CMF(2019)29/REV1), (FSB, 2020[51]). This, in turn, could give rise to concerns over potential risks to financial consumers, who may not be receiving the same range of product options, pricing or advice that would be provided through traditional financial services providers. It also could lead to difficulties for supervisors in accessing and auditing the financial activities provided by such firms.

Another related risk has to do with anti-competitive behaviours and market concentration in the technology aspect of the service provision. The possible emergence of a small number of key players in markets for AI solutions and/or services incorporating AI technologies (e.g. cloud computing service providers who also provide AI services), evidence of which is already observed in some parts of the world (ACPR, 2018[33]). Challenges for the competitive environment are also present given the privileged position BigTech players have with regards to customer data. In particular, such firms can use their data advantage to build monopolistic positions, both in relation to client acquisition (for example through effective price discrimination) and through the introduction of high barriers to entry for smaller players.

At the end of 2020, the European Union and the UK published regulatory proposals, the Digital Markets Act, that seek to establish an ex ante framework to govern ‘Gatekeeper’ digital platforms such as BigTech, aiming to mitigate some of the above risks and ensure fair and open digital markets (European Commission, 2020[52]). Some of the obligations proposed include the requirement for such Gatekeepers to provide business users with access to the data generated by their activities and provide data portability, while prohibiting them from using data obtained from business users to compete with these business users (to address dual role risks). The proposal also provides for solutions addressing self-referencing, parity

and ranking requirements to ensure no favourable treatment to the services offered by the Gatekeeper itself against those of third parties.

2.4. Integration of AI in Blockchain-based financial products

Applications of distributed ledger technologies (DLT), such as the blockchain, have proliferated in recent years across industries and primarily in finance. The rapid growth of blockchain-based applications is supported by the purported benefits of speed, efficiency and transparency that such innovative technologies may offer, driven by automation and disintermediation (OECD, 2020[53]). Widespread adoption of DLTs in finance may be driven by efforts to increase efficiencies from disintermediation, including in securities markets (issuance and post-trade/ clearing and settlement); payments (central bank digital currencies and fiat-backed stablecoins); and tokenisation of assets more broadly, and may drive the re-shaping of roles and business models of financial operators (e.g. custodians).

A convergence of AI and DLTs in blockchain-based finance is being promoted by the industry as a way to yield better results in such systems, as increased automation may amplify efficiencies promised by blockchain-based systems. However, the actual level of AI implementation in blockchain-based projects does not appear to be sufficiently large at this stage to justify claims of convergence between the two technologies.

Instead of a convergence, what is actually being observed in practice is the implementation of AI applications in certain blockchain systems, for specific use-cases (e.g. for risk management, see below), and similarly, the implementation of DLT solutions in certain AI mechanisms (e.g. for data management). The latter involves the use of DLTs to feed information to a ML model, making use of the immutable and disintermediated characteristics of the blockchain, while also allowing for the sharing of confidential information on a zero-knowledge basis without breaching confidentiality and privacy requirements. The use of DLTs in AI mechanisms is hypothetically expected to allow users of such systems to monetise data they own and which is being used by ML models and other AI-driven systems (e.g. IoT). Implementation of such AI use-cases is driven by the potential of the technology to further increase efficiency gains of automation and disintermediation in DLT-based systems and networks.

The largest contribution of AI in DLT-based finance may be in augmenting the automating capacity of smart contracts. Several applications of AI can be identified in specific use-cases applied within DLT networks, such as compliance and risk management (e.g. anti-fraud, introduction of automated restrictions to a network); and data inference and management (e.g. enhancing the function of Oracles[14]). Most of these applications are still in the development phase.

In particular, AI can be used in blockchain networks to reduce (but not eliminate) security susceptibilities and help protect against compromising of the network, for example in payment applications. Leveraging the power of AI can assist users of blockchain networks to identify irregular activities that could be associated with theft or scams, as such events do occur despite the need of both private and public keys in order to compromise security of a user. Similarly, AI applications can improve on-boarding processes on a network (e.g. biometrics for AI identification), as well as AML/CFT checks in the provision of any kind of DLT-based financial services. The integration of AI in DLT-based systems can also assist in compliance processes and risk management in such networks. For example, AI applications can provide a wallet address analysis results that can be used for regulatory compliance purposes or for an internal risk-based assessment of transaction parties (Ziqi Chen et al., 2020[54]). However, when financial intermediaries are eliminated from financial transactions, the effectiveness of the current financial regulatory approaches focusing on regulated entities may be eroded (Endo, 2019)[15].

The integration of AI-based solutions in DLT-based systems at the protocol level could help authorities achieve their regulatory objectives in an efficient manner. This could be achieved inter alia through the automatic sharing of regulated entities' data with the authorities in a seamless and real time manner, as

well as through the programming of regulatory requirements in the code of the programmes, promoting compliance in an automatic way. Participation of regulators as nodes in decentralised networks has been discussed by the market as one of the ways to resolve the challenges of supervision of such platforms that lack a single central authority.

When it comes to data, in theory, AI could be used in DLT-based systems to potentially improve the quality of the data inputs into the chain, as the responsibility of data curation shifts from third party nodes to independent, automated AI-powered systems, enhancing the robustness of information recording and sharing as such systems are more difficult to manipulate. In particular, the use of AI could improve the functioning of third party off-chain nodes, such as so-called 'Oracles', nodes feeding external data into the network. The use of Oracles in DLT networks carries the risk of erroneous or inadequate data feeds into the network by underperforming or even malicious third-party off-chain nodes (OECD, 2020[53]). In theory, the use of AI could further increase disintermediation by bringing AI inference directly on-chain, which would render third party providers of information to the chain, such as Oracles, redundant. In practice, it could act as a safeguard by testing the veracity and truthfulness of the data provided by the Oracles and prevent cyber-attacks or manipulation of such third-party data provision into the network.

The use of such AI applications could theoretically somehow increase participants' trust in the network as participants can test the information provided by the Oracle and check for any compromise in the system. It reality, however, the introduction of AI does not necessarily resolve the 'garbage in, garbage out' conundrum as the problem of poor quality or inadequate data inputs is a challenge that is also observed in AI-based mechanisms and applications (see Section 3.1).

2.4.1. AI augmenting the capabilities of smart contracts

The most significant impact from the integration of AI techniques in blockchain-based systems may come from their application in smart contracts, with a practical impact on the governance and risk management of such contracts and with numerous hypothetical (and yet untested) effects on roles and processes of DLT-based networks. In theory, the use of AI can allow for self-regulated DLT chains that will be operating on a fully autonomous basis.

Smart contracts have existed long before the advent of AI applications and rely on simple software code. Even today, most smart contracts used in a material way do not have ties to AI techniques. As such, many of the suggested benefits from the use of AI in DLT systems remains theoretical, and industry claims around convergence of AI and DLTs functionalities in marketed products should be treated with caution.

That said, AI use cases are extremely helpful in augmenting smart contract capabilities, particularly when it comes to risk management and the identification of flaws in the code of the smart contract. AI techniques such as NLP can be used to analyse the patterns of the smart contract execution and detect fraudulent activity and enhance the security of the system. More importantly, AI can perform testing of the code in a way that a human code reviewer cannot, in terms of both speed and level of detail/scenario analysis. Given that such code is the underlying basis for the automation of smart contracts, flawless coding is at the heart of the robustness of such contracts.

Box 2.4. Smart contracts in DLT-based systems

Smart contracts are distributed applications created and run over the blockchain, which consist of self-executing contracts written as code on Blockchain ledgers, automatically executed upon reaching pre-defined trigger events written in the code (OECD, 2019[55]).

Smart contracts are in essence programmes that run on the Ethereum blockchain. The coding of these programmes defines how they will operate and when. They define rules, like regular contracts, and automatically enforce them via the code once the conditions specified in the code are triggered.

Smart contracts are not controlled by a user but are deployed to the network and run as programmed. User accounts can then interact with a smart contract by submitting transactions that execute a function defined on the smart contract.

Smart contracts facilitate the disintermediation from which DLT-based networks can benefit, and are one of the major source of efficiencies that such networks promise to offer. They allow for the full automation of actions such as payments or transfer of assets upon triggering of certain conditions, which are pre-defined and registered in the code, without any human intervention. The legal status of smart contracts remains to be defined, as these are still not considered to be legal contracts in most jurisdictions (OECD, 2020[53]). Until it is clarified whether contract law applies to smart contracts, enforceability and financial protection issues will persist. The auditability of the code of such smart contracts also requires additional resources from market participants who will wish to confirm the basis on which such smart contracts are executed.

In theory, the deployment of AI in smart contracts could further enhance the automation capacity of smart contracts, by increasing their autonomy and allowing the underlying code to be dynamically adjusted according to market/environment conditions. The use of NLP, a subset of AI, could improve the analytical reach of smart contracts that are linked to traditional contracts, legislation and court decisions, going even further in analysing the intent of the parties involved (The Technolawgist, 2020[56]). It should be noted, however, that such applications of AI for smart contracts are purely theoretical at this stage and remain to be tested in real-life examples.

Challenges around operational risks as well as compatibility and interoperability of conventional infrastructure with DLT-based one and AI technologies remain to be examined. AI techniques such as deep learning require significant amounts of computational resources, which may pose an obstacle to performing well on the Blockchain (Hackernoon, 2020[57]). It has been argued that at this stage of development of the infrastructure, storing data off chain would be a better option for real time recommendation engines to prevent latency and reduce costs (Almasoud et al., 2020[58]). Operational risks associated with DLTs remain to be resolved as the technology and the applications facilitated by such technology matures.

2.4.2. Self-learning smart contracts and governance of DLTs: self-regulated chains and Decentralised Finance (DeFi)

In theory, AI-powered smart contracts can constitute the basis for the establishment of self-regulated chains. Researchers suggest that, in the future, AI could also be integrated for forecasting and automating in 'self-learned' smart contracts, similar to models applying reinforcement learning AI techniques (Almasoud et al., 2020[58]). In other words, AI can be used to extract and process information of real-time systems and feed such information into smart contracts. This means that code of smart contracts would

be adjusted automatically, and the governance of the chain would not require any human intervention, resulting in fully autonomous self-regulated decentralised chains.

Decentralised autonomous organisations (DAOs) are organisations that exist as autonomous code on the chain, have already existed, but could be further facilitated by AI-based techniques. For example, AI could provide real-time data feeds as inputs to the code, which in turn would calculate a desired action to take (Buterin, 2013[59]). AI-powered self-learning smart contracts would play a key role in adding new features in the logic of the chain, learn from the experience of the chain and adjust or introduce new rules, in essence defining the overall governance of the chain. Current DeFi projects are typically managed by DAOs, which governance has a variety of centralized aspects such as on-chain voting by governance token holders and off-chain consensus, and such human intervention could be a control point for regulators (Ushida and Angel, 2021[60]). However, the integration of AI into DAOs could facilitate further decentralization and reduce the enforceability of conventional regulatory approaches.

The use of AI to build fully autonomous chains raises important challenges and risks to its users and the wider ecosystem. In such environments, AI smart contracts, rather than humans, would execute decisions and operate systems without human intervention in the decision-making or operation of the system, with important ethical considerations arising from that. In addition, the introduction of automated mechanisms that switch off the model instantaneously (so-called 'kill switches') (see Section 2.2 for definitions) is very difficult in such networks, not least because of the decentralised nature of the network. This is one of the major issues that is also encountered in the DeFi space.

AI integration in blockchains could support decentralised applications in DeFi through use cases that increase automation and efficiencies in the provision of certain financial services. Indicatively, the introduction of AI models could support the provision of personalised/customised recommendations across products and services; credit scoring based on users' online data; investment advisory services and trading based on financial data; as well as other reinforcement learning[16] applications on blockchain-based processes (Ziqi Chen et al., 2020[54]). As in other blockchain-based financial applications, the deployment of AI in DeFi may augment the capabilities of the DLT use-case by providing additional functionalities, however, it may not radically affect any of the business models involved in DeFi applications.

Box 2.5. AI for ESG investing

ESG ratings vary strongly across ESG rating providers, due to different frameworks, measures, key indicators and metrics, data use, qualitative judgement, and weighting of subcategories (Boffo and Patalano, 2020[61]). Despite a proliferation of ESG ratings, market participants often lack the necessary tools (e.g. consistent data, comparable metrics, and transparent methodologies) to inform their decision-making in an appropriate manner (OECD, 2020[62]). The role of data is even more pertinent given the analysis of non-financial elements of company action and which relate to sustainability issues, but ESG data quality remains concerning due to gaps in data availability, potential inaccuracies and lack of comparability across providers.

AI and big data could be used for ESG investing to (i) assess company data (issuer data); (ii) assess non-company data; and (iii) assess the consistency and comparability of ratings to understand the drivers of scores. The purported benefit of AI is that it can allow for better informed decision-making by limiting the subjectivity and cognitive bias that may stem from traditional analysis, reducing the noise in ESG data and making use of unstructured data. In particular, NPL can be used to analyse massive amounts of unstructured datasets (geolocalisation, social media) in order to perform sentiment analysis, identify patters and relationships in these data. The results of such analysis can be used to assign quantitative values to qualitative data for sustainability parameters, based on AI techniques (Bala et al., 2014[63]).

Alternative ESG ratings providers have emerged, offering ratings based on AI with a view to provide a more objective, outside-in perspective of a companies' sustainability performance (Hughes, Urban and Wójcik, 2021[64]). The use of AI for ESG ratings generation could help overcome the risk of greenwashing by corporates that follow business-as usual strategies under the guise of sustainability, by uncovering less readily available information about the companies' practices and actions related to sustainability.

Empirical evidence of AI-driven alternative ESG ratings suggests that there are important advantages over traditional approaches, including higher levels of standardisation, a more democratic aggregation process, and the use of rigorous real-time analytics (Hughes, Urban and Wójcik, 2021[64]). Nevertheless, it is unlikely that such methods will replace traditional models in the future. Instead, they can complement traditional approaches to ESG rating, informing investors also aspects related to non-disclosed information of rated entities.

It should be noted, however, that the use of AI could also be used by rated corporates themselves, to further obscure the image of their sustainability action. AI techniques could allow such entities to better understand and quantitatively identify areas that should be strategically prioritised in terms of information disclosure so as to improve their ESG ratings, and emphasize them in order to manipulate their ESG ratings.

3 Emerging risks from the use of AI/ML/Big data and possible risk mitigation tools

As the use of AI/ML technology continues to grow in size and spectrum of applications in financial markets, a number of challenges and risks associated with such practices are being identified and may deserve further consideration by industry participants, users, and policy makers. Such challenges are apparent at many different levels: data level, model and firm level, social and systemic level.

This section examines some challenges arising in the use of AI-driven techniques in finance, which deserve further consideration as the application of AI in finance continues to grow, while it also touches on potential mitigants to such risks. Challenges discussed relate to data management and concentration; risk of bias and discrimination; explainability; robustness and resilience of AI models; governance and accountability in AI systems; regulatory considerations; employment risks and skills.

3.1. Data management

Data is at the core of any AI application, and the deployment of AI, ML models and big data offers opportunities for increased efficiencies, reduced costs, and greater customer satisfaction through the delivery of higher quality services/products.

This section examines how the use of big data in AI-powered applications could introduce an important source of non-financial risk to such financial products/services, driven by challenges and risks related to the quality of the data used; data privacy and confidentiality; cyber security; and fairness considerations. It discusses the risk of unintended bias and discrimination of parts of the population when data is misused or inappropriate data is being used by the model (e.g. in credit underwriting). It examines the importance of data when it comes to training, testing and validation of ML models, but also when defining the capacity of such models to retain their predictive powers in tail event situations. In addition to financial consumer protection considerations, there are potential competition issues arising from the use of big data and ML models, relating to potential high concentration amongst market providers. It should be noted that the challenges of data use and management identified and discussed below are not specific to big data/alternative date, but apply to data more broadly.

3.1.1. Representativeness and relevance of data

One of the four 'Vs' of big data, defined by the industry, is veracity, i.e. the uncertainty of the level of truthfulness of big data (IBM, 2020[11]). Such uncertainty may be stemming from doubtful source reliability, insufficient quality, or inadequate nature of the data used. With big data, veracity of observations may be affected by specific behaviours (e.g. social networks), noisy or biased data collection systems (e.g. sensors, IoT), and may prove insufficient to prevent or to mitigate disparate impact dynamics.

Data representativeness and relevance provide more precise attributes to data related to AI applications, as compared to data veracity. The former relates to whether data used provide an exhaustive representation of the population under study, with balanced representation of all relevant subpopulations. In financial markets, this could prevent over/under-representation of groups of operators, and enhance more accurate model training. In credit scoring, it could contribute to foster financial inclusion of minorities. Data relevance involves the contribution of data used to describe the phenomenon at hand without

including exogenous (misleading) information. For example, in credit scoring, relevance of information on natural persons' behaviour and/or reputation (for legal persons) should be carefully assessed prior to inclusion and usage by the model. Assessment of the dataset used on a case by case basis to improve accuracy and appropriateness of data used may be cumbersome given the sheer volume of data involved, while it may reduce the efficiencies delivered by the deployment of AI.

3.1.2. Data privacy and confidentiality

The volume, ubiquity and continuous flowing nature of data used in AI systems can raise various data protection and privacy concerns. In addition to standard concerns around the collection and use of personal data, potential incompatibilities arise in the area of AI, including through the power of AI to make inferences in big datasets; questionable feasibility of applying practices of 'notification and consent' that allow for privacy protection in ML models; as well as questions around data connectivity and the cross-border flow of data. The latter involves the importance of data connectivity in financial services and the critical importance of the ability to aggregate, store, process, and transmit data across borders for financial sector development, with the appropriate data governance safeguards and rules (Hardoon, 2020[65]).

The fusion of multiple datasets can present new big data users with new opportunities to aggregate data, while at the same time give rise to analytical challenges. Databases collected under heterogeneous conditions (i.e. different populations, regimes, or sampling methods) provide new opportunities for analysis that cannot be achieved through individual data sources. At the same time, the combination of such underlying heterogeneous environments gives rise to potential analytical challenges and pitfalls, including confounding, sampling selection, and cross-population biases (Bareinboim and Pearla, 2016[66]).

Cyber security risks, risk of hacking and other operational risks witnessed across the board of digital financial products/services have direct implications on data privacy and confidentiality. While the deployment of AI does not open up possibilities of new cyber breaches, it could exacerbate pre-existing ones by, *inter alia*, linking falsified data and cyber breaches, creating new attacks which can alter the functioning of the algorithm through the introduction of falsified data into models or the alteration of existing ones (ACPR, 2018[33]).

Consumers' financial and non-financial data are increasingly being shared and used, sometimes without their understanding and informed consent (US Treasury, 2018[8]). While informed consent is the legal basis for any use of data, consumers are not necessarily educated on how their data is handled and where it is being used, and consent may not be well informed. Increased tracking of online activity with advanced modes of tracking increases such risks, and so does data sharing by third party providers. Observed data not provided by the customer, such as geolocation data or credit card transaction data are prime examples of datasets at risk of possible violations of privacy policy and data protection laws.

New approaches are being suggested by the industry to secure non-disclosive computation, which protects consumer privacy, including through the generation and use of tailor-made, synthetic datasets, which are put together for the purposes of ML modelling, or the use of Privacy Enhancing Technologies (PETs). PETs try to preserve the overall properties and characteristics of the original data without revealing information about actual individual data samples. PETs include differential privacy, federated analysis, homomorphic encryption and secure multi-party computation. Particularly, differential privacy provides mathematical guarantees on the desired level of privacy and allows better accuracy compared to synthetic datasets. The purported advantage of such techniques is that models trained on synthetic data instead of real data do not show a significant loss of performance. In terms of ensuring the privacy data handling in models, data anonymization approaches do not provide rigorous privacy guarantees, especially given inferences made by AI-based models.

The use of big data by AI-powered models could expand the universe of data that is considered sensitive, as such models can become highly proficient in identifying users individually (US Treasury, 2018[8]). Facial recognition technology and other inferred data such as customer profile can be used by the model to

identify users or infer other characteristics, such as gender, when joined up with other information. AI models could achieve the re-identification of anonymised databases by cross-referencing publicly available databases and narrowing down matches to ultimately attribute sensitive information to individuals (Luminovo.ai, 2020[67]). What is more, the higher dimensionality in ML data sets, i.e. the possibility to take into account an unlimited number of variables compared to conventional statistical techniques, increases the likelihood of sensitive information being included in the analysis.

Regulators have renewed their focus on data privacy and protection driven by increased digitalisation of the economy (e.g. EU GDPR) and aiming to reinforce consumer protection across markets, rebalance the power relationship of corporates and individuals, shifting power back to the consumers, and ultimately increase transparency and trust in how companies use consumer data. 'Protection of Consumer Data and Privacy' is one of the Principles of the G20/OECD High-Level Principles on Financial Consumer Protection (OECD, 2011[68]). Protection of individuals' personal data in the use of AI in finance is also at the core of the Monetary Authority of Singapore's principles to promote fairness, ethics, accountability and transparency (MAS, 2019[69]).

From the industry standpoint, one of the stumbling blocks in better data governance for financial sector firms relates to the perceived fragmentation in regulatory and supervisory responsibility around data, and to which institutions would be accountable for applying best practices of data governance in areas such as data quality, definitions, standardisation, architecture, deduplications, and other. Such fragmentation is magnified in case of cross-border activities.

The economics of data use are being redefined in parallel with the rapid deployment of ML models in finance. A small number of alternative dataset players have emerged, exploiting the surge in demand for datasets that inform AI techniques, with limited visibility and overseeing over their activity at this stage. The purchase and use of datasets by such small niche database providers is possibly raising risks around their lawful purchase and use by financial service providers. Increased compliance costs of regulations aiming to protect consumers may further redefine the economics of the use of big data for financial market providers and, by consequence, their approach in the use of AI and big data.

3.2. Data concentration and competition in AI-enabled financial services/products

The strength and nature of the competitive advantages created by advances in AI could potentially harm the operations of efficient and competitive markets if consumers' ability to make informed decisions is constrained by high concentrations amongst market providers (US Treasury, 2018[8]). To the extent that the deployment of AI and proprietary models provides a performance edge against competition, this may, in turn, result in restricted participation by smaller financial services providers who may not have the financial resources and human capital necessary for them to adopt in-house AI/ML techniques or use big data information sources. Unequal access to data and potential dominance in the sourcing of big data by few big BigTech in particular, could reduce the capacity of smaller players to compete in the market for AI-based products/services.

The potential for network effects further amplifies the risks of concentration and dependencies on few large players, which could in turn result in the emergence of new systemically important players. BigTech are the prime example of such potential risk, and the fact that they fall outside the regulatory perimeter adds to the challenges involved. This is mainly driven by the access and use of data by BigTech and gets amplified through the use of AI techniques to monetise such data. Increasingly, a small number of alternative data providers are shaping the economics of database provision, with some risk of concentration developing in that market.

In terms of data-driven barriers to entry in the AI market, smaller firms may face disproportionate burdens in the deployment of such technologies, given the requirement for expensive complementary assets such as advanced data-mining and ML software, as well as physical infrastructures such as data centres, whose

investment is subject to economies of scale. The ability of algorithms to find new relations and patterns of behaviour also requires access to a variety of data collected from multiple sources, resulting in economies of scope. Thereby, small firms that do not have the necessary complementary assets or that are not simultaneously present in multiple markets might face barriers to entry, preventing them from developing algorithms that can effectively exert competitive pressure (OECD, 2016a).

Healthy competition in the market for AI-based financial products/services is vital for providers to be able to fully unleash the benefits of the technology, particularly when it comes to trading and investing. The use of outsourced/ third party vendor models could 'arbitrage away' the benefits of such tools for firms adopting them and could result in one-way markets and herding behaviour by financial consumers or convergence of trading/investment strategies by finance practitioners.

3.2.1. Risk of tacit collusions

The widespread deployment of AI-based models could also raise competition issues by making tacit collusion[17] easier without any formal agreement or human interaction (OECD, 2017[35]). In a tacitly collusive context, the non-competitive outcome is achieved by each participant deciding its own profit-maximising strategy independently of its competitors (OECD, 2017[35]).[18] In other words, the use of algorithms makes it easier for market participants to sustain profits above the competitive level without having entered into an agreement, effectively replacing explicit collusion with tacit co-ordination.

Even though tacit collusion typically occurs in transparent markets with few market participants, there is evidence that collusion might become easier to sustain and more likely to be observed when algorithms are involved in digital markets characterised by high transparency and frequent interaction (OECD, 2017[35]).

The dynamic adaptive capacity of self-learning and deep learning AI models can therefore raise the risk that the model recognises the mutual interdependencies and adapts to their behaviour and actions of other market participants or other AI models, possibly reaching a collusive outcome without any human intervention and perhaps without even being aware of it (OECD, 2017[35]). Although such collusions are not necessarily illegal from a competition law standpoint, questions are raised on whether and how enforcement action could be applied to the model and its users in case that is the case.

3.3. Risk of bias and discrimination

Depending on how they are used, AI methods have the potential to help avoid discrimination based on human interactions, or intensify biases, unfair treatment and discrimination in financial services. By delegating the human-driven part of the decision-making to the algorithm, the user of the AI-powered model avoids biases attached to human judgement. At the same time, the use of AI applications may risk bias or discrimination through the potential to compound existing biases found in the data; by training models with such biased data; or through the identification of spurious correlations (US Treasury, 2018[8]).

The use of flawed or inadequate data may result in wrong or biased decision-making in AI systems. Poor quality data can result in biased or discriminatory decision-making through two avenues. ML models trained with inadequate data risk producing inaccurate results even when fed with good quality data. Equally, ML models that are trained on high-quality data will certainly produce a questionable output if they are then fed with unsuitable data, despite the well-trained underlying algorithm. Well-intentioned ML models may inadvertently generate biased conclusions, discriminated against protected classes of people (White & Case, 2017[46]). The use of incorrect, inaccurate (e.g. poorly labelled, incomplete) or even fraudulent data in ML models carries the risk of 'garbage in, garbage out', and the quality of the data determines largely the quality of the model's output.

Biases may also be inherent in the data used as variables and, given that the model trains itself on data from external sources that may have already incorporated certain biases, perpetuates historical biases. In

addition, biased or discriminatory decisions made by ML models are not necessarily intentional and can even occur with strong quality, well-labelled data, through inference and proxies, or given the fact that correlations between sensitive and 'non-sensitive' variables may be difficult to detect in vast databases (Goodman and Flaxman, 2016[70]). As big data involves vast amounts of data reflecting society, AI-driven models could just perpetuate biases that already exist in society and are reflected in such databases.

Box 3.1. Labelling and structuring of data used in ML models

Labelling and structuring of data is an important, albeit tedious task, necessary for ML models to be able to perform. AI can only distinguish the signal from the noise if it can successfully identify and recognise what a signal is, and models need well-labelled data to be able to recognise patterns in them (S&P, 2019[48]). To that end, supervised learning models (the most common form of AI) require feeding software stacks of pre-tagged examples classified in a consistent manner, until the model can learn to identify the data category by itself.

That said, the correct data labels might not be readily apparent from simple observation data points. Data labelling is a labour-intensive process that requires analysis of vast amounts of data and which is currently reported to be outsourced to specialised firms or to the distributed workforce (The Economist, 2019[71]). Analysis and labelling of data by humans present opportunities to identify errors and biases in the data used, although according to some it may inadvertently introduce other biases as it involves subjective decision-making.

As the process of data, cleansing and labelling could be prone to human error, and a number of solutions involving AI themselves have started to develop. Considerations around the quality of the data and its level or representativeness can help avoid unintended biases at the output level.

In addition, and given the high dimensionality of data, users of ML models need to adequately identify those features of the data that are relevant to the scenarios tested by the model. Different methods are being developed to reduce the existence of irrelevant features or 'noise' in datasets and improve ML model performance. An interesting alternative is the use of artificial or 'synthetic' datasets generated and employed for this purpose, as well as for testing and validation purposes (see Section 3.5).

Note: These considerations do not apply to unsupervised learning models, which identify patterns in data that have not been labelled by humans.
Source: (The Economist, 2019[71]) (S&P, 2019[48]), (Calders and Verwer, 2010[72]).

The role of humans in decision-making processes informed by AI is critical in in identifying and correcting for biases built into the data or in the model design. Also, in order to explain the output of the model, although the extent to which all this is feasible remains an open question (US Treasury, 2018[8]). The human parameter is critical both at the data input stage and at the query input stage and a degree of scepticism in the evaluation of the model results can be critical in minimising the risks of biased model output/decision-making.

The design of a ML model and its audit can further strengthen the degree of assurance about the robustness of the model when it comes to avoiding potential biases. Inadequately designed and controlled AI/ML models carry a risk of exacerbating or reinforcing existing biases while at the same time making discrimination even harder to observe (Klein, 2020[73]). Auditing mechanisms of the model and the algorithm that sense check the results of the model against baseline datasets can help ensure that there is no unfair treatment or discrimination by the technology (see Section 3.4.1). Ideally, users and supervisors

should be able to test scoring systems to ensure their fairness and accuracy (Citron and Pasquale, 2014[50]). Tests can also be run based on whether protected classes can be inferred from other attributes in the data, and a number of techniques can be applied to identify and/or rectify discrimination in ML models (Feldman et al., 2015[74]). Governance of AI/ML models and the assignment of accountability to the human parameter of the project is also important to safeguard prospective borrowers against possible unfair biases. When measuring for potential biases, it may be important to avoid comparing ML-based decisioning to a hypothetical unbiased state but use realistic reference points, comparing such methods to traditional statistical models and human-based decision-making, both of which are not flawless or utterly unbiased.

3.4. Explainability

Perhaps the most widely acknowledged challenge of ML models is the difficulty in decomposing the output of a ML model into the underlying drivers of its decision, in other words understanding why and how the model generates results. This difficulty in justifying or rationalising model decisions and outputs is generally described by the term 'explainability'. AI-based models are inherently complex given the nature of the technology employed. The possible intentional concealment by market players of the mechanics of AI models, in order to protect their intellectual property, is reinforcing such lack of explainability. Given the generalised gap in technical literacy, for most end-user consumers having access to underlying code is insufficient to explain the mechanics of the model. This issue is aggravated by the mismatch between the complexity characterising AI models and the demands of human-scale reasoning or styles of interpretation that fit the human cognition (Burrell, 2016[75]).

The perceived lack of trust of users and supervisors around AI applications can be largely attributed to the lack of explainability of ML models. AI-powered approaches in finance are becoming increasingly opaque, and even if the underlying mathematical principles of such models can be explained, they still lack 'explicit declarative knowledge' (Holzinger, 2018[76]). Improving the explainability levels of AI applications can therefore contribute to maintaining the level of trust by financial consumers and regulators/supervisors, particularly in critical financial services (FSB, 2017[77]). From an internal control and governance perspective, a minimum level of explainability needs to be ensured for a model committee to be able to analyse the model brought to the committee and be comfortable with its deployment.

The lack of explainability could be incompatible with existing regulation where this requires the interpretation of the underlying logic and/or the reporting of such logic. For example, regulation may require algorithms to be fully understood and explainable throughout their lifecycle (IOSCO, 2020[78]). Other policies may grant citizens a 'right to explanation' for decisions made by algorithms and information on the logic involved, such as the GDPR in the EU[19] applied in credit decisions or insurance pricing, for instance. Another example is the potential use of ML in the calculation of regulatory requirements (e.g. risk-weighted assets or RWA for credit risk), where the existing rules require that the model be explainable or at least subject to human oversight and judgement (e.g. Basel Framework for Calculation of RWA for credit risk – Use of models 36.33).[20]

The lack of explainability in ML-based models used by financial market participants could become a macro-level risk if not appropriately supervised by micro prudential supervisors, as it becomes difficult for both firms and supervisors to predict how models will affect markets (FSB, 2017[79]). In particular, AI could introduce or amplify systemic risks related to pro-cyclicality, given increased risk of herding and convergence of strategies by users of 'off-the-shelf' third party provider models. In the absence of an understanding of the detailed mechanics underlying a model, users have limited room to predict how their models affect market conditions, and whether they contribute to market shocks. Users are also unable to adjust their strategies in time of poor performance or in times of stress, leading to potential episodes of exacerbated market volatility and bouts of illiquidity during periods of acute stress, aggravating flash crash

type of events. Risks of market manipulation (e.g. spoofing, see section 2.2) or tacit collusions (see section 3.2.1) are also present in the absence of an understanding of the underlying mechanics to the model.

Financial market practitioners using AI-powered models are facing greater scrutiny over the explainability of their models. Due in part to heightened attention, many market participants are working to improve the explainability of such models so as to be able to better comprehend their behaviour in normal market conditions and in times of stress, and to manage associated risks. Contrary to post hoc explainability of a single decision, explainability by design, i.e. incorporated into the AI mechanism, is more difficult to achieve given that (i) the audience may be unable to grasp the logic of the model; (ii) some models are by definition impossible to fully comprehend (e.g. some neural networks); and (iii) the full revealing of the mechanism is equivalent to giving away IP.

An interesting debate associated with explainability concerns the question of whether and how AI explainability should be any different to that required in the application of other complex mathematical models in finance. There is a risk that AI applications are held to a higher standard and thus subjected to a more onerous explainability requirement as compared to other technologies, with negative repercussions for innovation (Hardoon, 2020[80]). The objective of the explainability analysis at committee level should focus on the underlying risks that the model might be exposing the firm to, and whether these are manageable, instead of its underlying mathematical promise.

Given the trade-off between explainability and performance of the model, financial services providers need to strike the right balance between explainability of the model and accuracy/performance. Some degree of insight into the workings and the underlying logic of the model, as well as the reasoning followed in its decision-making, prevent models from being considered as 'black boxes' and may allow firms to comply with regulatory requirements, while also building trust with consumers. Some jurisdictions do not accept any models that are complete black boxes and where there is no degree of explainability achieved (e.g. Germany).

It should also be highlighted that there is no need for a single principle or one-size-fits-all approach for explaining ML models, and explainability will depend to a large extent on the context (Brainard Lael, 2020[81]) (Hardoon, 2020[80]). When looking into the interpretability of the model, one needs to take into account who is asking the question and what the model is predicting. In addition, ensuring the explainability of the model does not by itself guarantee that the model is reliable (Brainard Lael, 2020[81]). Contextual alignment of explainability with the audience needs to be coupled with a shift of the focus towards 'explainability of the risk', i.e. understanding the resulting risk exposure from the use of the model instead of the methodology underlying such model. Recent guidance issued by the UK Information Commissioner's Office suggests using five contextual factors to help in assessing the type of explanation needed: domain, impact, data used, urgency, and audience (see Box 4.3) (UK Information Commissioner's Office, 2020[82]).

Box 3.2. Explaining decisions made with AI: Guidance by the UK Information Commissioner's Office

The UK Information Commissioner's Office has issued guidance on information provision around AI-based decision making, including five contextual factors affecting why people want explanations of such decisions.

These contextual factors include:

- domain - the setting or sector of activity;
- impact - the effect of the decision on the individual;
- data - data used to train and test the model, which is likely to influence the consumer's willingness to accept or contest an AI-based decision;
- urgency: how much time the consumer has to reflect on the decision; and,
- audience: which individuals is the firm explaining an AI-driven decision to, which defines what type of information is meaningful/useful to them.

Guidance was also provided on the prioritisation of explanations of AI-assisted decisions, and emphasised the importance of developing understanding and knowledge of AI use among the general public.

Source: (UK Information Commissioner's Office, 2020[82]).

3.4.1. Auditability of AI algorithms and models

The underlying complexity of such 'black box' models raises regulatory challenges in the transparency and auditing of such models in many financial services use cases (e.g. lending) (US Treasury, 2018[8]). It is difficult, if not impossible, to perform an audit on a ML model if one cannot decompose the outcome of the model into its underlying drivers. Lack of explainability does not allow the supervisor to follow the process that led to a model outcome, limiting the possibility of auditing. A number of laws or regulations in some jurisdictions have been designed around a baseline expectation of auditability and transparency, which may not be easily achievable when AI-enabled models are used. Audit trails can only be followed if one can provide evidence of the sequence of activities or processes, and this capacity is curtailed by the lack of interpretability of some AI models. As decisions made by such models no longer follow a linear process, and the models themselves are characterised by limited interpretability, there is a need to find ways to improve the explainability of AI outcomes, while ensuring accountability and robust governance dynamics in AI-based systems.

Research efforts that aim at improving the interpretability of AI-driven applications and rendering ML models more amenable to ex ante and ex post inspection are underway both in academia (Vellido, Martín-Guerrero and Lisboa, 2012[83]) as well as by the industry.

3.4.2. Disclosure

Based on the OECD AI Principles, 'there should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them' (OECD, 2019[5]). It is argued that the opacity of algorithm-based systems could be addressed through transparency requirements, ensuring that clear information is provided as to the AI system's capabilities and limitations (European Commission, 2020[84]). Separate disclosure should inform consumers about the use of AI system in the delivery of a product and their interaction with an AI system instead of a human being (e.g.

robo-advisors). Such disclosure can also allow customers to make conscious choices among competing products.

To date, there is no commonly accepted practice as to the level of disclosure that should be provided to investors and financial consumers and potential proportionality in such information. According to market regulators, there should be a differentiation as to the level of transparency depending on the type of investor (retail vs. institutional), as well as the area of implementation (front vs. back office) (IOSCO, 2020[78]). Suitability requirements, such as the ones applicable to the sale of investment products, might help firms better assess whether the prospective clients have a solid understanding of how the use of AI affects the delivery of the product/service.

Requirements for financial firms to document in writing operational details and design characteristics of models used in finance were already in place before the advent of AI. Documentation of the logic behind the algorithm, to the extent feasible, is being used by some regulators as a way to ensure that the outcomes produced by the model are explainable, traceable and repeatable (FSRA, 2019[85]). The EU, for instance, is considering requirements around disclosure documentation of programming and training methodologies, processes and techniques used to build, test, and validate AI systems, including documentation on the algorithm (what the model shall optimise for, which weights are designed to certain parameters at the outset etc.) (European Commission, 2020[84]). The US Public Policy Council of the Association for Computing Machinery (USACM) has proposed a set of principles targeting inter alia transparency and auditability in the use of algorithms, suggesting that models, data, algos and decisions be recorded so as to be available for audit where harm is suspected (ACM US Public Policy Council, 2017[86]). The Federal Reserve's guidance for model risk management includes also documentation of model development and validation that is sufficiently detailed to allow parties unfamiliar with a model to understand how the model operates, as well as its limitations and key assumptions (Federal Reserve, 2011[87]).

Financial service providers find it harder to document the model process of AI-enabled models used for supervisory purposes (Bank of England and FCA, 2020[88]). Challenges in explaining how the model works translate into difficulties in documenting such complex models, irrespective of the size of the service provider. Some jurisdictions have proposed a two-pronged approach to AI model supervision: (i) analytical: combining analysis of the source code and of the data with methods (if possible based on standards) for documenting AI algorithms, predictive models and datasets; and (ii) empirical: leveraging methods providing explanations for an individual decision or for the overall algorithm's behaviour, and relying on two techniques for testing an algorithm as a black box: challenger models (to compare against the model under test) and benchmarking datasets, both curated by the auditor (ACPR, 2020[89]).

In addition to explainability-related challenges, AI-based models require the setting of a wealth of parameters that have a significant effect on model performance and results. Such parameterisation could be considered as 'arbitrary' and subjective, as they may be based on intuition rather than validation and they can be very much dependant on the designer of the model. Transparency around the parameters selected in the model could potentially address part of the issue, as the way the model works with these parameters remains largely difficult to explain.

3.5. Robustness and resilience of AI models: training and testing performance

AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed (OECD, 2019[5]). The robustness of AI systems can be reinforced by careful training of models, as well as testing of the performance of models based on their intended purpose.

3.5.1. Training AI models, validating them and testing their performance

In order to capture higher order interactions (i.e. non-linearity), models may need to be trained with a larger size of datasets as higher order effects are harder to find. As such, the datasets used for training must be large enough to capture non-linear relationships and tail events in the data. This is hard to achieve in practice, given that tail events are rare and the dataset may not be robust enough for optimal outcomes. At the same time, using ever-larger sets of data for training models risks making models static, which, in turn, may reduce the performance of the model and its ability to learn.

The inability of the industry to train models on datasets that include tail events is creating a significant vulnerability for the financial system, weakening the reliability of such models in times of unpredicted crisis and rendering AI a tool that can be used only when market conditions are stable. ML models carry a risk of over-fitting, when a trained model performs extremely well on the samples used for training but performs poorly on new unknown samples, i.e. the model does not generalise well (Xu and Goodacre, 2018[90]). To mitigate this risk, modellers split the data into training and test/validation set and use the training set to build the (supervised) model with multiple model parameter settings; and the test/validation set to challenge the trained model, assess the accuracy of the model and optimise its parameters. The validation set contains samples with known provenance, but these classifications are not known to model, therefore, predictions on the validation set allow the operator to assess model accuracy. Based on the errors on the validation set, the optimal model parameters set is determined using the one with the lowest validation error (Xu and Goodacre, 2018[90]).

The measured performance of validation models was previously considered by scientists as an unbiased estimator of the performance of such models, however, multiple recent studies have demonstrated that this assumption does not always hold (Westerhuis et al., 2008[91]), (Harrington, 2018[92]). As highlighted by such studies, having an additional blind test set of data that is not used during the model selection and validation process is necessary to have a better estimation of the generalisation performance of the model. Such validation processes go beyond the simple back testing of a model using historical data to examine ex-post its predictive capabilities, and ensure that the model's outcomes are reproducible.

Synthetic datasets are being artificially generated to serve as test sets for validation, and provide an interesting alternative given that they can provide inexhaustible amounts of simulated data, and a potentially cheaper way of improving the predictive power and enhancing the robustness of ML models, especially where real data is scarce and expensive. Some regulators require, in some instances, the evaluation of the results produced by AI models in test scenarios set by the supervisory authorities (e.g. Germany) (IOSCO, 2020[78]).

Ongoing monitoring and validation of models throughout their life is fundamental for the risk management of any type of model (Federal Reserve, 2011[87]) (see Box 3.4). Model validation is carried out after model training and helps confirm that the model is appropriately implemented, as well as that the model is being used and performing as intended. It comprises a set of processes and activities intended to verify that models are performing in line with their design objectives and business uses, while ensuring that models are sound. This is achieved through the identification of potential limitations and assumptions, and the assessment of possible impact. All model components, including input, processing, and reporting, should be subject to validation, and this applies equally to models developed in-house and to those outsourced or provided by third parties (Federal Reserve, 2011[93]). Validation activities should be performed on an ongoing basis to track known model limitations and identify any new ones, especially during periods of stressed economic or financial conditions, which may not be reflected in the training set.

Continuous testing of ML models is indispensable in order to identify and correct for 'model drifts' in the form of concept drifts or data drifts. Concept drifts are old concepts (Widmer, 1996[94]) and describe situations where the statistical properties of the target variable studied by the model change, which changes the very concept of what the model is trying to predict. For example, the definition of fraud could evolve over time with new ways of conducting illegal activity, such a change would result in concept drift.

Data drifts occur when statistical properties of the input data change, affecting the model's predictive power. The major shift of consumer attitudes and preferences towards ecommerce and digital banking is a good example of such data drifts not captured by the initial dataset on which the model was trained and result in performance degradation.

Ongoing monitoring and validation of ML models is a very effective way to prevent and address such drifts, and standardised procedures for such monitoring may assist in improving model resilience, and identify whether the model necessitates adjustment, redevelopment, or replacement. Related to this, it is of key importance to have an effective architecture in place that allows models to be rapidly retrained with new data as data distribution changes, so as to mitigate risks of model drifts.

Box 3.3. Guidance for model risk management in the US and EU that applies to AI models

Supervision and regulatory letter SR 11-7 issued by the Federal Reserve in 2011 provides technology-neutral guidance on model risk management that has stood the test of time, and is certainly useful in managing risks related to AI-driven models (Federal Reserve, 2011[87]).

The letter provides guidance on model development, implementation and use by banking institutions, and looks into (i) model development, implementation, and use; (ii) model validation and use; and (iii) governance, policies and controls.

More recently, the European Banking Authority (EBA) published Guidelines on loan origination and monitoring, including rules for appropriate management of model risks. The EBA aims to ensure that such guidelines are both future proof and technology neutral (EBA, 2020[95]).

In addition to ongoing monitoring and reviewing of the code/model used, some regulators have imposed the existence of 'kill switches' or other automatic control mechanisms that trigger alerts under high risk circumstances. Kill switches are an example of such control mechanisms that can quickly shut down an AI-based system in case it ceases to function according to the intended purpose. In Canada, for instance, firms are required to have built-in 'override' functionalities that automatically disengage the operation of the system or allows the firm to do so remotely, should need be (IIROC, 2012[96]). Such kill switches need to be tested and monitored themselves, to ensure that firms can rely on them in case of need.

There may be a need to reinforce existing risk management functions and processes related to models, in order to reflect emerging risks and/or unintended consequences related to the use of AI-based models. For example, the performance of models may need to be tested in extreme market conditions, to prevent systemic risks and vulnerabilities that may arise in times of stress. The data used to train the model may not fully reflect stressed market conditions or changes in exposures, activities or behaviours, therefore creating model limitations and likely deteriorating model performance. The recent use of such models also means that they remain untested at addressing risk under shifting financial conditions. It is therefore important to use a range of scenarios for testing and back-testing to allow for consideration of shifts in market behaviour and other trends, hopefully reducing the potential for underestimating risk in such scenarios (FSB, 2017[77]).

Interestingly, research suggests that explainability that is 'human-meaningful' can significantly affect the users' perception of a system's accuracy, independent of the actual accuracy observed (Nourani et al., 2020[97]). When less human-meaningful explanations are provided, the accuracy of the technique that does not operate on human-understandable rationale is less likely to be accurately judged by the users.

3.5.2. Correlation without causation and meaningless learning

The intersection of causal inference and ML is a rapidly expanding area of research (Cloudera, 2020[98]). The understanding of cause-and-effect relationships is a key element of human intelligence that is absent

from pattern recognition systems. Researchers in deep learning are increasingly recognising the importance of such questions, and using them to inform their research, although such type of research is still at a very early stage.

Users of ML models could risk interpreting meaningless correlations observed from patterns in activity as causal relationships, resulting in questionable model outputs. Moving from correlation to causation is crucial when it comes to understanding the conditions under which a model may fail, as it will allow us to understand whether we can expect the pattern to continue to be predictive over time. Causal inference is also instrumental when it comes to the ability to replicate empirical findings of a model to new environments, settings or populations (i.e. external validity of the model's output). The ability to transfer causal effects learned in the test dataset to a new set of data, in which only observational studies can be conducted, is called transportability and is fundamental for the usefulness and robustness of ML models (Pearl and Bareinboim, 2014[99]). It may be useful for supervisors to have some understanding about the casual assumptions AI model users make in order to better assess potential associated risks.

Outputs of ML models need to be evaluated appropriately and the role of human judgement is fundamental to that end, especially when it comes to the question of causation. When not interpreted with a certain degree of scepticism or caution, correlation without causation identified in patterns by AI-based models may result in biased or false decision-making. Evidence from some research suggests that models are bounded to learn suboptimal policies if they do not take into account human advice, perhaps surprisingly, even when human's decisions are less accurate than their own (Zhang and Bareinboim, 2020[100]).

3.5.3. AI and tail risk: the example of the COVID-19 crisis

Although AI models are adaptive in nature, as they evolve over time by learning from new data, they may not be able to perform under idiosyncratic one-time events that have not been experienced before, such as the COVID-19 crisis, and which are therefore not reflected in the data used to train the model. As AI-managed trading systems are based on dynamic models trained on long time series, they are expected to be effective in as long as the market environment has some consistency with the past. Evidence based on a survey conducted in UK banks suggest that around 35% of banks experienced a negative impact on ML model performance during the pandemic (Bholat, Gharbawi and Thew, 2020[101]). This is likely because the pandemic has created major movements in macroeconomic variables, such as rising unemployment and mortgage forbearance, which required ML (as well as traditional) models to be recalibrated.

Tail and unforeseen events, such as the recent pandemic, give rise to discontinuity in the datasets, which in turn creates model drifts that undermine the models' predictive capacity (see Section 3.5.1). Tail events cause unexpected changes in the behaviour of the target variable that the model is looking to predict, and previously undocumented changes to the data structure and underlying patterns of the dataset used by the model, both caused by a shift in market dynamics during such events. These are naturally not captured by the initial dataset on which the model was trained and are likely to result in performance degradation. Synthetic datasets generated to train the models could going forward incorporate tail events of the same nature, in addition to data from the COVID-19 period, with a view to retrain and redeploy redundant models.

Ongoing testing of models with validation datasets that incorporate extreme scenarios and continuous monitoring for model drifts is therefore of paramount importance to mitigate risks encountered in times of stress. It should be noted that models based on reinforcement learning, where the model is trained in simulated conditions, are expected to perform better in times of one-off tail risk events, as they are easier to train based on scenarios of extreme unexpected market conditions that may not have been observed in the past.

3.6. Governance of AI systems and accountability

Solid governance arrangements and clear accountability mechanisms are fundamentally important as AI models are deployed in high-value decision-making use-cases (e.g. in determining who gets access to credit or how investment portfolio allocation is decided). Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning (OECD, 2019[5]). In addition, human oversight from the product design and throughout the lifecycle of the AI products and systems may be needed as a safeguard (European Commission, 2020[84]).

Currently, financial market participants using AI rely on existing governance and oversight arrangements for the use of such technologies, as AI-based algorithms are not considered to be fundamentally different from conventional ones (IOSCO, 2020[78]). Existing governance frameworks applicable to models can constitute the basis for the development or adaptation for AI activity, given that many of the considerations and risks associated with AI are also applicable to other types of models. Explicit governance frameworks that designate clear lines of responsibility for the development and overseeing of AI-based systems throughout their lifecycle, from development to deployment, could further strengthen existing arrangements for operations related to AI. Internal governance frameworks could include minimum standards or best practice guidelines and approaches for the implementation of such guidelines (Bank of England and FCA, 2020[88]). Internal model committees set model governance standards and processes that financial service providers are following for model building, documentation, and validation for any time of model, including AI-driven ML models (see Box 3.5).

Model governance frameworks have yet to address how to handle AI models, which exist only ephemerally, and change very frequently. The challenge of using existing model governance processes for ML models is associated with more advance AI models that rebuild themselves at relatively short time intervals. One possible mitigating approach would be to retain data and code so that replications of the inputs and outputs can be produced based on a past date. However, since many ML models are non-deterministic, there is no guarantee that even with the same input data the same model will be produced.

Importantly, intended outcomes for consumers would need to be incorporated in any governance framework, together with an assessment of whether and how such outcomes are reached using AI technologies. In advanced deep learning models, issues may rise concerning the ultimate control of the model, as AI could unintentionally behave in a way that is contrary to consumer interests (e.g. biased results in credit underwriting, described above). In addition, the autonomous behaviour of some AI systems during their life cycle may entail important product changes having an impact on safety, which may require a new risk assessment (European Commission, 2020[84]).

Box 3.4. ML model governance and model committees

In general, financial service providers use the same model governance processes for model building, documentation, and model validation for ML models as for traditional statistical models.

Model governance best practices were introduced since the emergence of the use of traditional statistical models for credit and other consumer finance decisions. In particular, financial institutions must ensure that models are built using appropriate datasets; that certain data is not used in the models; that data that is a proxy for a protected class is not used; that models are rigorously tested and validated (sometimes by independent validators); and that when models are used in production, the production input data is consistent with the data used to build the model. Documentation and audit trails are also held around for example deployment decisions, design, production.

Model governance frameworks also provide that models must be monitored to ensure they do not produce results that constitute comparative evidence of disparate treatment. Importantly, it must be possible to determine why the model produced a given output.

Model governance committees or model review boards are in place in financial services firms with the aim of designing, approving and overseeing the implementation of model governance processes. Model validation is part of such processes, using holdout datasets. Other standard processes include the monitoring for stability in inputs, outputs and parameters. Such internal committees are expected to become more common with the wider adoption of AI by financial firms, with possible 'upgrading' of their roles and competencies and some of the processes involved to accommodate for the complexities introduced by AI-based models. For example, the frequency of model validation, and the validation methods for AI-based models need to be different to the ones applying to linear models.

AI is also being deployed for RegTech purposes, and as part of their model governance, financial services companies are making efforts to enhance the automated processes that monitor and control the data that is consumed by the models in production, as well as to enhance the automated monitoring of model outputs.

Ultimate responsibility, oversight and accountability over AI-based systems lies by definition with Executive and board level of management of the financial services provider, who have to establish an organisation-wide approach to model risk management, and ensure that the level of model risk is within their tolerance. At the same time, the importance of other functions such as engineers/programmers or data analysts, who have previously not been central to the supervisors' review, may end up subject to increased scrutiny, commensurate with their increased importance in the deployment of AI-based financial products/services. Accountabilities for AI-related systems would therefore perhaps need to go beyond senior managers and board to hands-on professionals responsible for programming and development of models and to those using the mechanism to deliver customer services, at a minimum at the internal risk management level, as responsibility of explanation of such models to senior managers and the Board lies with the technical functions. Interestingly, some jurisdictions may require a third party audit to validate the performance of the model in line with the intended purpose (FSRA, 2019[85]). Strong governance also includes documentation of model development and validation (see section 3.4.2.).

3.6.1. Outsourcing and third party providers

Risks arise also when it comes to outsourcing of AI techniques to third parties, both in terms of competitive dynamics (concentration risks) and in terms of giving rising to systemic vulnerabilities related to increased risk of convergence.

Possible risks of concentration of certain third party providers may rise either in terms of data collection and management (e.g. dataset providers) or in the area of technology (e.g. third party model providers) and infrastructure (e.g. cloud providers) provision. AI models and techniques are being commoditised through cloud adoption, and the risk of dependency on providers of outsourced solutions raises new challenges for competitive dynamics and potential oligopolistic market structures in such services.

The use of third-party models could also give rise to risks of convergence at the firm level and the systemic level, especially if there is a lack of heterogeneity of third-party models in the market. This could translate into herding and bouts of illiquidity in times of stress, when liquidity is most needed. Such impact is likely to be further amplified by the reduced warehousing capacity of traditional market-makers, who would otherwise stabilise markets by providing ample liquidity during periods of market stress through active market making. The risk of herding is more prominent in smaller institutions, which are most likely to rely on third party providers to outsource ML model creation and management, as they lack the in-house expertise to fully understand and govern such models.

The outsourcing of AI techniques or enabling technologies and infrastructure raises challenges in terms of accountability in addition to concentration risks. Governance arrangements and contractual modalities are important in managing risks related to outsourcing, similar to those applying in any other type of services. Finance providers need to have the skills necessary to audit and perform due diligence over the services provided by third parties. Over-reliance in outsourcing may also give rise to increased risk of disruption of service with potential systemic impact in the markets. Contingency and security plans need to be in place to allow business to function as usual if any vulnerability materialises.

3.7. Regulatory considerations, fragmentation and potential incompatibility with existing regulatory requirements

Although many countries have dedicated AI strategies (OECD, 2019[5]), a very small number of jurisdictions have current requirements that are specifically targeting AI-based algorithms and models. In most cases, regulation and supervision of ML applications are based on overarching requirements for systems and controls (IOSCO, 2020[78]). These consist primarily of rigorous testing of the algorithms used before they are deployed in the market, and continuous monitoring of their performance throughout their lifecycle.

The technology-neutral approach that is being applied by most jurisdictions to regulate financial market products (in relation to risk management, governance, and controls over the use of algorithms) may be challenged by the rising complexity of some innovative use-cases in finance. Given the depth of technological advances in AI areas such as deep learning, existing financial sector regulatory regimes could fall short in addressing the systemic risks posed by a potential broad adoption of such techniques in finance (Gensler and Bailey, 2020[102]).

What is more, some advanced AI techniques may not be compatible with existing legal or regulatory requirements. The lack of transparency and explainability of some ML models and the dynamic nature of continuously adapting deep learning models are prime examples of such potential incompatibility. Inconsistencies may also arise in areas such as data collection and management: the EU GDPR framework for data protection imposes time constraints in the storing of individual data, but AI-related rules could require firms to keep record of datasets used to train the algorithms for audit purposes. Given the sheer size of such datasets, there are also practical implications and costs involved in the recording of data used to train models for supervisory purposes.

Some jurisdictions, such as the EU, have identified a possible need to adjust or clarify existing legislation in certain areas (e.g. liability) in order to ensure an effective application and enforcement (European Commission, 2020[84]). This comes because of the opaqueness of AI systems, which makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation. In the medium term, regulators and supervisors may need to adjust regulations and supervisory methods to adapt to new realities introduced by the deployment of AI (e.g. concentration, outsourcing) (ACPR, 2018[33]).

Industry participants note a potential risk of fragmentation of the regulatory landscape with respect to AI at the national, international and sectoral level, and the need for more consistency to ensure that these techniques can function across borders (Bank of England and FCA, 2020[88]). In addition to existing regulation that is applicable to AI models and systems, a multitude of published AI principles, guidance, and best practice have been developed in recent years. While these are all seen by the industry as valuable in addressing potential risks, views differ over their practical value and the difficulty of translating such principles into effective practical guidance (e.g. through real life examples) (Bank of England and FCA, 2020[88]).

The ease of use of standardised, off-the-shelf AI tools may encourage non-regulated entities to provide investment advisory or other services without proper certification/licensing in a non-compliant way. Such regulatory arbitrage is also observed with mainly BigTech entities making use of datasets they have access to from their primary activity.

3.8. Employment risks and the question of skills

Financial services providers and supervisors need to be technically capable of operating, inspecting AI-based systems and intervening when required. The absence of adequate skills is a potential source of vulnerabilities for both the industry side and the regulatory/supervisory side, and which may give rise to potential employment issues in the financial industry. The deployment of AI and big data in finance requires different skillsets that a relatively small segment of financial practitioners possess. In line with significant investments that will need to be made to develop AI-based models and tools, firms will need to also develop human capital with the requisite skills to derive value from such technologies and exploit value from vast amounts of unstructured data sources.

From an industry viewpoint, the deployment of AI involves the use of professionals who combine scientific expertise in the area of AI, computer science (programming, coding) skills and financial sector expertise. While today's financial market participants have somehow silo-ed roles for specialists in IT or finance, the widespread use of AI by financial institutions will increasingly rely on, and drive demand for, experts who successfully combine finance knowledge with expertise in computer science. Importantly, compliance professionals and risk managers will need to have the appropriate understanding of the workings of the AI techniques and models to be able to audit, oversee, challenge and approve their use. Similarly, senior managers, who in most cases are accountable for the use of such techniques, need to be able to understand and follow their development and implementation.

The widespread adoption of AI and ML by the financial industry may give rise to some employment challenges. On the one hand, demand for employees with applicable skills in AI methods, advanced mathematics, software engineering and data science is expected to be significant. On the other hand, executives of financial services firms expect that the application of such technologies may result in potentially significant job losses across the industry (Noonan, 1998[103]) (US Treasury, 2018[8]). In practice, it is expected that financial market practitioners and risk management experts will gain experience and expertise in AI in the medium term, as AI models will coexist with traditional models and until the time AI become more mainstream.

Over-reliance in fully automated AI-based systems may give rise to increased risk of disruption of service with potential systemic impact in the markets. If markets dependent on such systems face technical or other disruptions, financial service providers need to ensure that from a human resources perspective, they are ready to substitute the automated AI systems with well-trained humans acting as a human safety net and capable of ensuring there is no disruption in the markets. Such considerations are likely to become increasingly important, as the deployment of AI becomes ubiquitous across markets.

The issue of skills and technical expertise becomes increasingly important also from a regulatory and supervisory perspective. Financial sector regulators and supervisors may need to keep pace with the technology and enhance the skills necessary to effectively supervise AI-based applications in finance. Enforcement authorities may need to be technically capable of inspecting AI-based systems and empowered to intervene when required (European Commission, 2020[84]). The upskilling of policy makers will also allow them to expand their own use of AI in RegTech and SupTech, an important area of application of innovation in the official sector (see Section 1.2.2).

AI in finance should be seen as a technology that augments human capabilities instead of replacing them. It could be argued that a combination of 'man and machine', where AI informs human judgment rather than replaces it (decision *aid* instead of decision *maker*), could allow for the benefits of the technology to realise, while maintaining safeguards of accountability and control as to the ultimate decision-making. At the current stage of maturity of AI solutions, and in order to ensure that vulnerabilities and risks arising from the use of AI-driven techniques are minimised, some level of human supervision of AI-techniques is still necessary. The identification of converging points, where human and AI are integrated, will be critical for the practical implementation of such a combined 'man and machine' approach ('human in the loop').

4 Policy responses and implications

4.1. Recent policy activity around AI and finance

Given the potentially transformative effect of AI on certain markets, as well as the new types of risks stemming from its use, AI has been a growing policy priority for the past few years. In May 2019, the OECD adopted its Principles on AI, the first international standards agreed by governments for the responsible stewardship of trustworthy AI, with guidance from a multi-stakeholder expert group. The issues addressed through the OECD AI Principles, and the link of the Principles to sustainable and inclusive growth make them highly relevant for application to global finance.

Box 4.1. The OECD AI Principles

The OECD Council adopted the Recommendation on AI at Ministerial level on 22-23 May 2019. The OECD AI Principles focus on how governments can shape a human-centric approach to trustworthy AI, and aim to promote the use of AI that is innovative and trustworthy, respecting human rights and democratic values.

The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI:

- AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

The OECD also provides five recommendations to governments:

- Facilitate public and private investment in research & development to spur innovation in trustworthy AI.
- Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.
- Ensure a policy environment that will open the way to deployment of trustworthy AI systems.
- Empower people with the skills for AI and support workers for a fair transition.
- Co-operate across borders and sectors to progress on responsible stewardship of trustworthy AI.

Source: (OECD, 2019[5]).

In 2020, the European Commission issued a White Paper with policy and regulatory options for an 'AI ecosystem for excellence and trust' (European Commission, 2020[84]). The proposal outlines specific actions for the support, development and uptake of AI across the EU economy and public administration; provides options for a future regulatory framework on AI; and discusses safety and liability aspects on AI. Action at European level is also directed at the practical implementation level, with initiatives such as the Infinitech consortium's pilot projects financed by the EC and aiming at lowering the barriers for AI- driven innovation, boosting regulatory compliance and stimulating investments in the sector (Infinitech, 2020[104]) (see Box 4.2).

Box 4.2. Project Infinitech: testing AI applications in finance at the European level

Project Infinitech is a consortium-led initiative comprising 48 participants in 16 EU member countries, funded by the European Commission's Horizon 2020 Research and Innovation Programme and aiming to test more than 20 pilots and testbeds in the areas of digital finance, leveraging the benefits of AI, big data, and IoT.

Infinitech's pilot AI-driven products and services include use-cases around Know Your Customer (KYC), customer analytics, personalised portfolio management, credit risk assessment, financial crime and fraud prevention, insurance and RegTech tools incorporating data governance capabilities and facilitating compliance to regulations (e.g. PSD2, 4AMLD, MIFiD II).

Some examples of Infinitech pilot projects include:

- Smart and Automated Credit Risk Assessment for SMEs: A big data platform integrating an AI and Blockchain-based system for credit risk scoring of SMEs;
- Real-time risk assessment in Investment Banking: real time risk assessment and monitoring procedure for two standard risk metrics – VaR (Value-at-Risk) and ES (Expected Shortfall);
- Collaborative Customer-centric Data Analytics for Financial Services: An AI-based support tools for new customer services, consisting of a system based on data sharing, a credit scoring system and an AML system based on semantic technologies and DLT-based data sharing;
- Personalized Portfolio Management: AI-based portfolio construction for wealth management regardless of portfolio size;
- Platform for Anti Money Laundering (AML) Supervision: Aiming to improve the effectiveness of the existing supervisory activities (analysis reports, risk assessment and screening tool) by processing Big Data;
- Real-time cybersecurity analytics on Financial Transactions' Big Data: real-time analysis of financial transactions of mobile banking, applying ML models in combination with analytics techniques on high-volume data streams, enabling proper and prompt countermeasures to anomalies.

Source: (Infinitech, 2020[104]), (Poljšak Borut, Bank of Slovenia).

In 2019, the IOSCO board identified AI and ML as an important priority. In 2020, IOSCO published a consultation report on the use of AI by market intermediaries and asset managers, proposes six measures to assist IOSCO members in creating appropriate regulatory frameworks to supervise market intermediaries and asset managers that use such technologies (see Box 4.3).

Box 4.3. IOSCO's consultation on the use of AI and ML by market intermediaries and asset managers

In June 2020, IOSCO published a consultation on the use of AI and ML by market intermediaries and asset managers, to assist its members in providing appropriate regulatory frameworks in the supervision of market intermediaries and asset managers that use AI and ML (IOSCO, 2020[78]).

As part of the consultation, IOSCO proposes guidance consisting of six measures that reflect expected standards of conduct, to ensure that the following features are secured:

- Appropriate governance, controls and oversight frameworks over the development, testing, use and performance monitoring of AI and ML;
- Staff have adequate knowledge, skills and experience to implement, oversee, and challenge the outcomes of the AI and ML;
- Robust, consistent and clearly defined development and testing processes to enable firms to identify potential issues prior to full deployment of AI and ML; and
- Appropriate transparency and disclosures to investors, regulators and other relevant stakeholders.

Source: (IOSCO, 2020[78]).

Efforts have also been made at the national level. In 2018, the French ACPR established a taskforce bringing together professionals from the financial industry (business associations, banks, insurers, FinTechs) and public authorities to discuss current and potential uses of AI in the industry, the associated opportunities and risks, as well as the challenges faced by supervisors (ACPR, 2018[33]). In 2019, the Bank of England and the Financial Conduct Authority launched the AI Public Private Forum (see Box 4.4). The Russian Federation enacted a National Strategy for the development of AI in 2019, and a Concept for the development of regulation for AI technologies and robotics in 2020. In 2021, a Federal Law on Experimental Digital Innovation Regimes came into force, empowering Bank of Russia to approve the launch of regulatory sandboxes, including for projects deploying AI solutions in finance. A five-year regulatory sandbox for implementing AI has been launched already in Moscow in July 2020 under a special Federal Law.

In terms of more recent policy and regulatory initiatives, on 31 March 2021, the Comptroller of the Currency, the Federal Reserve System, the Federal Deposit Insurance Corporation, the Consumer Financial Protection Bureau, and the National Credit Union Administration issued a Request for Information and comment on financial institutions' use of AI, including ML (Federal Register, 2021[105]). The consultation notes the benefits and main risks of AI in finance (around explainability, data usage, and dynamic updating) and seeks input on questions related to explainability, the broader or more intensive data processing and usage, risk of overfitting, cybersecurity risks, fair lending considerations, oversight of third parties and other considerations (Federal Register, 2021[105]).

On 21 April 2021, the European Commission published a proposal for a regulation that aims to address the risks of AI and lay down harmonised rules on the use of AI across sectors of activity, while also proposes the establishment of a European AI Board (European Commission, 2021[106]). While the proposal's overall scope is wide, the strongest requirements apply to the high-risk applications of AI, which includes assessment of creditworthiness. The obligations for such high-risk AI include requirements to use detailed and specific risk and quality management systems and subject the system to a conformity assessment; use high-quality data that is representative, free from errors and complete; keep records and logs, and be transparent to users about the use and operation of the AI-driven applications. The proposed

rules also introduce a requirement for human oversight by suitably trained individuals, the use of kill switches and/or explicit human confirmation of decision making; ensure the accuracy, robustness and security of the system; conduct post-market monitoring and notify the regulator about serious incidents, as well as register the system on a public register.

4.2. Policy considerations

The increased deployment of AI in financial services can provide important benefits to financial consumers and market participants, by improving the quality of services offered and producing efficiencies to financial service providers. At the same time, AI-based applications in finance can give rise to new challenges (e.g. related to lack of explainability) or amplify risks that already exist in financial markets (e.g. related to data management and use).

Policy makers and regulators have a role in ensuring that the use of AI in finance is consistent with promoting financial stability, protecting financial consumers, and promoting market integrity and competition. Emerging risks from the deployment of AI techniques need to be identified and mitigated to support and promote the use of responsible AI. Existing regulatory and supervisory requirements may need to be clarified and sometimes adjusted in order to address some of the perceived incompatibilities of existing arrangements with AI applications.

The application of regulatory and supervisory requirements on AI techniques could be looked at under a contextual and proportional framework, depending on the criticality of the application and the potential impact on the consumer involved. This will likely encourage the use of AI without unnecessarily stifling innovation.

Policy makers should consider sharpening their focus on better data governance by financial sector firms, aiming to reinforce consumer protection across AI applications in finance. Some of the most important risks raised in this note relate to data management: data privacy, confidentiality, concentration of data and possible impact on the competitive dynamics of the market, but also risk of unintended bias and discrimination of parts of the population and data drifts. The importance of data is undisputed when it comes to training, testing and validation of ML models, but also when defining the capacity of such models to retain their predictive powers in tail event situations.

Policy makers could consider the introduction of specific requirements or best practices for data management in AI-based techniques. These could touch upon data quality, adequacy of the dataset used depending on the intended use of the AI model, and safeguards that provide assurance about the robustness of the model when it comes to avoiding potential biases. Appropriate sense checking of model results against baseline datasets and other tests based on whether protected classes can be inferred from other attributes in the data are two examples of best practices to mitigate risks of discrimination. The validation of the appropriateness of variables used by the model could reduce a source of potential biases. Tools could be developed and used to monitor and correct for conceptual drifts. Requirements for additional transparency over the use of personal data and opt-out options for the use of personal data could be considered by authorities.

Policy makers could consider disclosure requirements around the use of AI techniques in the provision of financial services and that may impact the customer outcome. Financial consumers need to be informed about the use of AI techniques in the delivery of a product, as well as potential interaction with an AI system instead of a human being, in order to be able to make conscious choices among competing products. Clear information around the AI system's capabilities and limitations should be included in such disclosure. The introduction of suitability requirements for AI-driven financial services, similar to the ones applicable to the sale of investment products, could be considered by authorities. Such requirements would help financial service providers better assess whether prospective clients have a solid understanding of how the use of AI affects the delivery of the product.

The limited transparency and explainability of many advanced AI-based ML models is a key policy question that remains to be resolved. Lack of explainability is incompatible with existing laws and regulations, but also with internal governance, risk management and control frameworks of financial service providers. It limits the ability of users to understand how their models affect markets or contributes to market shocks. It can amplify systemic risks related to pro-cyclicality, convergence, and increased market volatility through simultaneous purchases and sales of large quantities, particularly when third party standardised models are used. Importantly, the inability of users to adjust their strategies in times of stress may lead to exacerbated market volatility and bouts of illiquidity during periods of acute stress, aggravating flash crash type of events.

Regulators should consider how to overcome the perceived incompatibility of the lack of explainability in AI with existing laws and regulations. There may be a need to update and/or adjust the currently applicable frameworks for model governance and risk management by financial services firms in order to address such challenges arising by the use of AI-based models. The supervisory focus may need to be shifted from documentation of the development process and the process by which the model arrives to its prediction to model behaviour and outcomes, and supervisors may wish to look into more technical ways of managing risk, such as adversarial model stress testing or outcome-based metrics (Gensler and Bailey, 2020[2]).

Despite recent progress to improve the explainability of AI from low levels, explainability remains at the core of the perceived lack of trust of users and supervisors around AI applications. While current discussions tend to focus on improving explainability as the sole mechanism to promote trust, other checks and balances may need to be introduced to ensure that ML model-based decisioning is operating as intended.

Policy makers could consider requiring clear model governance frameworks and attribution of accountability to the human in order to help build trust in AI-driven systems. Explicit governance frameworks that designate clear lines of responsibility for the development and overseeing of AI-based systems throughout their lifecycle, from development to deployment, may need to be put in place by financial services providers so as to strengthen existing arrangements for operations related to AI. Internal model governance frameworks may need to be adjusted to better capture risks emerging from the use of AI, as well as to incorporate intended outcomes for consumers together with an assessment of whether and how such outcomes are reached using AI technologies. Adequate documentation and audit trails of the above processes can assist the oversight of such activity by supervisors.

The provision of increased assurance by financial firms around the robustness and resilience of AI models is fundamental as policy makers seek to guard against build-up of systemic risks, and will help AI applications in finance gain trust. The performance of models may need to be tested in extreme market conditions, to prevent systemic risks and vulnerabilities that may arise in times of stress. The introduction of automatic control mechanisms (such as kill switches) that trigger alerts or switch off models in times of stress could assist in mitigating risks, although they expose the firm to new operational risks. Back-up plans, models and processes should be in place to ensure business continuity in case the models fails or acts in unexpected ways. Further, regulators could consider add-on or minimum buffers if banks were to determine risk weights or capital based on AI algorithms (Gensler and Bailey, 2020[2]).

Frameworks for appropriate training, retraining and rigorous testing of AI models may need to be introduced and/or reinforced to ensure that ML model-based decisioning is operating as intended and in compliance with applicable rules and regulations. Datasets used for training must be large enough to capture non-linear relationships and tail events in the data, even if synthetic, to improve the reliability of such models in times of unpredicted crisis. Continuous testing of ML models is indispensable in order to identify and correct for model drifts.

The ongoing monitoring and validation of AI models, which are fundamental for their risk, should be further promoted by regulators, as the most effective way to improve model resilience and prevent and address model drifts. Best practices around standardised procedures for such monitoring and validation could

assist in improving model resilience, and identify whether the model necessitates adjustment, redevelopment, or replacement. Model validation, and the necessary approvals and sign-offs would need to be separated from the development of the model and documented as best possible for supervisory purposes. The frequency of testing and validation may need to be defined depending on the complexity of the model and the materiality of the decisions made by such model.

Appropriate emphasis could be placed on human primacy in decision making when it comes to higher-value use-cases (e.g. lending decisions) which significantly affect consumers. Authorities could consider the introduction of processes that can allow customers to challenge the outcome of AI models and seek redress could also help build trust over such systems. The GDPR is an example of such policies, as it provides the associated right of individuals 'to obtain human intervention' and to express their points of view if they wish to contest the decision made by an algorithm (EU, 2016[3]). Public communication by the official sector that clearly sets expectations can further build confidence in AI applications in finance.

Policy makers should consider the increased technical complexity of AI, and whether resources will need to be deployed to keep pace with advances in technology. Investment in research can allow some of the issues around explainability and unintended consequences of AI techniques to be resolved. Investment in skills for both finance sector participants and policy makers will allow them to follow advancements in technology and maintain a multidisciplinary dialogue at operational, regulatory and supervisory level. Closer cooperation of IT staff with more traditional finance experts could be one way to adjust the trade-off between predictability of the model and explainability and respond to the legal and regulatory requirements for auditability and transparency. There may be a need to build bridges between disciplines that currently work in silos, such as deep learning and symbolic approaches (with the latter involving rules created through human intervention), to help improve explainability in AI-based approaches (European Commission, 2020[84]). Enforcement authorities in particular may need to be technically capable of inspecting AI-based systems and empowered to intervene when required, but also to enjoy the benefits of this technology by deploying AI in RegTech/ SupTech applications.

The role of policy makers is important in supporting innovation in the sector while ensuring that financial consumers and investors are duly protected and the markets around such products and services remain fair, orderly and transparent. Policy makers may need to sharpen their existing arsenal of defences against risks emerging from, or exacerbated by, the use of AI. Clear communication around the adoption of AI and the safeguards in place to protect the system and its users can help instil trust and confidence and promote the adoption of such innovative techniques. Given the ease of cross-border provision of financial services, a multidisciplinary dialogue between policy makers and the industry should be fostered and maintained both at national and international levels.

References

(n.a.) (n.d.), *Infinitech - The project*, https://www.infinitech-h2020.eu/the-project (accessed on 1 March 2021). [119]

ACM US Public Policy Council (2017), *Principles for Algorithmic Transparency and Accountability*. [86]

ACPR (2020), *Governance of Artificial Intelligence in Finance*, https://acpr.banque-france.fr/sites/default/files/medias/documents/20200612_ai_governance_finance.pdf (accessed on 6 April 2021). [89]

ACPR (2018), *Artificial intelligence: challenges for the financial sector*. [33]

Albanesi, S. and D. Vamossy (2019), "Predicting Consumer Default: A Deep Learning Approach", *SSRN Electronic Journal*, http://arxiv.org/abs/1908.11498 (accessed on 8 June 2021). [40]

Almasoud, A. et al. (2020), *Toward a self-learned Smart Contracts*. [58]

Bala, G. et al. (2014), *Tracking Companies' Real Time Sustainability Trends: Cognitive Computing's Identification of Short-Term Materiality Indicators 1 [working title] Draft DO NOT QUOTE OR CITE WITHOUT PERMISSION (comments appreciated)*, http://www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/Documents/building- (accessed on 8 June 2021). [63]

Bank of England and FCA (2020), *Minutes: Artificial Intelligence Public-Private Forum-First meeting*. [88]

Bank of Italy (2019), *Corporate default forecasting with machine learning*. [39]

BarclayHedge (2018), *BarclayHedge Survey: Majority of Hedge Fund Pros Use AI/Machine Learning in Investment Strategies.*, https://www.barclayhedge.com/insider/barclayhedge-survey-majority-of-hedge-fund-pros-use-ai-machine-learning-in-investment-strategies (accessed on 3 December 2020). [17]

Barclays Investment Bank (2020), *Algo Wheels*, https://www.investmentbank.barclays.com/our-insights/algo-wheels-6-building-blocks.html (accessed on 23 December 2020). [27]

Bareinboim, E. and J. Pearla (2016), "Causal inference and the data-fusion problem", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 113/27, pp. 7345-7352, http://dx.doi.org/10.1073/pnas.1510507113. [66]

Bholat, D., M. Gharbawi and O. Thew (2020), *The impact of Covid on machine learning and data science in UK banking | Bank of England*, Bank of England Quarterly Bulletin Q4 2020, https://www.bankofengland.co.uk/quarterly-bulletin/2020/2020-q4/the-impact-of-covid-on-machine-learning-and-data-science-in-uk-banking (accessed on 15 March 2021). [101]

BIS (2020), *Data vs collateral*, http://www.bis.org (accessed on 1 September 2020). [42]

BIS (2019), *How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm*, http://www.bis.org (accessed on 2 September 2020). [114]

BIS Markets Committee (2020), *Markets Committee FX execution algorithms and market functioning*, http://www.bis.org (accessed on 7 June 2021). [32]

Blackrock (2019), *Artificial intelligence and machine learning in asset management Background*. [14]

Bloomberg (2019), *What's an "Algo Wheel?" And why should you care? | Bloomberg Professional Services*, https://www.bloomberg.com/professional/blog/whats-algo-wheel-care/ (accessed on 14 December 2020). [25]

Bluwstein, K. et al. (2020), *Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach | Bank of England*, https://www.bankofengland.co.uk/working-paper/2020/credit-growth-the-yield-curve-and-financial-crisis-prediction-evidence-from-a-machine-learning (accessed on 15 March 2021). [23]

BNY Mellon (2019), *Artificial Intelligence Sweeps Hedge Funds*, https://www.bnymellon.com/us/en/insights/all-insights/artificial-intelligence-sweeps-hedge-funds.html (accessed on 3 December 2020). [20]

Boffo, R. and R. Patalano (2020), *ESG Investing: Practices, Progress and Challenges*. [61]

Brainard Lael (2020), *Speech by Governor Brainard on supporting responsible use of AI and equitable outcomes in financial services - Federal Reserve Board*, https://www.federalreserve.gov/newsevents/speech/brainard20210112a.htm (accessed on 12 January 2021). [81]

Brookings (2020), *Reducing bias in AI-based financial services*, https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/ (accessed on 27 August 2020). [47]

Burrell, J. (2016), "How the machine 'thinks': Understanding opacity in machine learning algorithms", http://dx.doi.org/10.1177/2053951715622512. [75]

Buterin, V. (2013), *Bootstrapping A Decentralized Autonomous Corporation: Part I*, https://bitcoinmagazine.com/articles/bootstrapping-a-decentralized-autonomous-corporation-part-i-1379644274 (accessed on 12 January 2021). [59]

Calders, T. and S. Verwer (2010), *Three naive Bayes approaches for discrimination-free classification*, Springer, http://dx.doi.org/10.1007/s10618-010-0190-x. [72]

CFA (n.d.), *AI, What Have You Done for Us Lately? | CFA Institute Enterprising Investor*, 2020, https://blogs.cfainstitute.org/investor/2019/10/14/ai-what-have-you-done-for-us-lately/ (accessed on 14 December 2020). [21]

Ching TM (2020), *Artificial intelligence: New cyber risks, new security capabilities | DXC Technology*, https://www.dxc.technology/security/insights/148677-artificial_intelligence_new_cyber_risks_new_security_capabilities (accessed on 8 June 2021). [34]

Citron, D. and F. Pasquale (2014), "The Scored Society: Due Process for Automated Predictions", https://papers.ssrn.com/abstract=2376209 (accessed on 2 September 2020). [50]

Cloudera (2020), *Causality for Machine Learning*, https://ff13.fastforwardlabs.com/ (accessed on 19 March 2021). [98]

Deloitte (2019), *Artificial intelligence The next frontier for investment management firms*. [15]

Dobbie, W. et al. (2018), *Measuring bias in consumer lending*, http://www.nber.org/papers/w24953 (accessed on 8 June 2021). [44]

EBA (2020), *Guidelines on loan origination and monitoring*, https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Guidelines/2020/Guidelines%20on%20loan%20origination%20and%20monitoring/884283/EBA%20GL%202020%2006%20Final%20Report%20on%20GL%20on%20loan%20origination%20and%20monitoring.pdf (accessed on 18 March 2021). [95]

Equifax (2017), *Equifax Announces Cybersecurity Incident Involving Consumer Information*, https://investor.equifax.com/news-and-events/press-releases/2017/09-07-2017-213000628 (accessed on 2 September 2020). [110]

EU (2016), *EUR-Lex - 32016R0679 - EN - EUR-Lex*, https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 11 March 2021). [3]

European Commission (2021), *Proposal for a Regulation laying down harmonised rules on artificial intelligence | Shaping Europe's digital future*, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (accessed on 9 June 2021). [106]

European Commission (2020), *On Artificial Intelligence-A European approach to excellence and trust White Paper on Artificial Intelligence A European approach to excellence and trust*, https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf. (accessed on 11 December 2020). [84]

European Commission (2020), *The Digital Services Act package | Shaping Europe's digital future*, https://ec.europa.eu/digital-single-market/en/digital-services-act-package (accessed on 24 March 2021). [52]

Experian (2018), *The State of Alternative Credit Data*. [108]

Federal Register (2021), *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning*, https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence (accessed on 9 June 2021). [105]

Federal Register (n.d.), *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning*, 2021, https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence (accessed on 9 June 2021). [120]

Federal Reserve (2011), *SUPERVISORY GUIDANCE ON MODEL RISK MANAGEMENT CONTENTS*. [93]

Federal Reserve (2011), *The Fed - Supervisory Letter SR 11-7 on guidance on Model Risk Management -- April 4, 2011*, https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm (accessed on 18 March 2021). [87]

Feldman, M. et al. (2015), *Certifying and removing disparate impact*, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/2783258.2783311. [74]

Financial Stability Board (2017), *Artificial intelligence and machine learning in financial services Market developments and financial stability implications*, http://www.fsb.org/emailalert (accessed on 27 August 2020). [36]

Financial Times (2020), *Hedge funds: no market for small firms | Financial Times*, https://www.ft.com/content/d94760ec-56c4-4051-965d-1fe2b35e4d71 (accessed on 3 December 2020). [18]

FinReg Lab (2019), *The Use of Cash-Flow Data in Underwriting Credit*, http://www.flourishventures.com (accessed on 27 August 2020). [109]

FSB (2020), *BigTech Firms in Finance in Emerging Market and Developing Economies*, http://www.fsb.org/emailalert (accessed on 12 January 2021). [51]

FSB (2020), *The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market developments and financial stability implications*, http://www.fsb.org/emailalert (accessed on 18 March 2021). [12]

FSB (2017), *Artificial intelligence and machine learning in financial services Market developments and financial stability implications*, http://www.fsb.org/emailalert (accessed on 1 December 2020). [77]

FSB (2017), *Artificial Intelligence and Machine Learning In Financial Services Market Developments and Financial Stability Implications*, http://www.fsb.org/emailalert (accessed on 1 December 2020). [79]

FSRA (2019), *Supplementary Guidance-Authorisation of Digital Investment Management ("Robo-advisory") Activities*. [85]

Fuster, A. et al. (2017), "Predictably Unequal? The Effects of Machine Learning on Credit Markets", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3072038. [43]

Gensler, G. and L. Bailey (2020), "Deep Learning and Financial Stability", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3723132. [102]

Gensler, G. and L. Bailey (2020), "Deep Learning and Financial Stability", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3723132. [2]

Goodman, B. and S. Flaxman (2016), *European Union regulations on algorithmic decision-making and a ``right to explanation''*. [70]

Gould, M. (2016), *Why the Finance Industry is Ripe for AI Disruption - Techonomy*, https://techonomy.com/2016/09/why-the-finance-industry-is-ripe-for-ai-disruption/ (accessed on 10 December 2020). [16]

GPFI (2017), *Alternative Data Transforming SME Finance*, http://www.istock.com (accessed on 2 September 2020). [107]

Gregoriou, G. and N. Duffy (2006), "Hedge funds: A summary of the literature", *Pensions: An International Journal*, Vol. 12/1, pp. 24-32, http://dx.doi.org/10.1057/palgrave.pm.5950042. [115]

Hackermoon.com (2020), *Why AI + Blockchain Make Sense? | Hacker Noon*, https://hackernoon.com/why-ai-blockchain-make-sense-5k4u3s6l (accessed on 12 January 2021). [9]

Hackernoon (2020), *Running Artificial Intelligence on the Blockchain | Hacker Noon*, https://hackernoon.com/running-artificial-intelligence-on-the-blockchain-77490d37e616 (accessed on 12 January 2021). [57]

Hardoon, D. (2020), *Contextual Explainability*, https://davidroihardoon.com/blog/f/contextual-explainability?blogcategory=Explainability (accessed on 15 March 2021). [80]

Hardoon, D. (2020), *Welcome Remarks for Address by Under Secretary McIntosh*, https://davidroihardoon.com/blog/f/welcome-remarks-for-address-by-under-secretary-mcintosh?blogcategory=Data (accessed on 15 March 2021). [65]

Harrington, P. (2018), *Multiple Versus Single Set Validation of Multivariate Models to Avoid Mistakes*, Taylor and Francis Ltd., http://dx.doi.org/10.1080/10408347.2017.1361314. [92]

Holzinger, A. (2018), "From Machine Learning to Explainable AI", https://www.researchgate.net/profile/Andreas_Holzinger/publication/328309811_From_Machine_Learning_to_Explainable_AI/links/5c3cd032a6fdccd6b5ac71e6/From-Machine-Learning-to-Explainable-AI.pdf (accessed on 2 September 2020). [76]

Hughes, A., M. Urban and D. Wójcik (2021), "Alternative ESG Ratings: How Technological Innovation Is Reshaping Sustainable Investment", *Sustainability*, Vol. 13/6, pp. 1-23, https://ideas.repec.org/a/gam/jsusta/v13y2021i6p3551-d522385.html (accessed on 8 June 2021). [64]

Hurley, M. (2017), *Credit scoring in the era of big data*, Yale Journal of Law and Technology. [49]

IBM (2020), *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? | IBM*, https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks (accessed on 31 May 2021). [31]

IBM (2020), *The Four V's of Big Data | IBM Big Data & Analytics Hub*, https://www.ibmbigdatahub.com/infographic/four-vs-big-data (accessed on 1 December 2020). [11]

IDC (2020), *Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching $110 Billion in 2024, According to New IDC Spending Guide*, https://www.idc.com/getdoc.jsp?containerId=prUS46794720 (accessed on 31 May 2021). [1]

IIROC (2012), *Rules Notice Guidance Note Guidance Respecting Electronic Trading*. [96]

Infinitech (2020), *Infinitech - The Flagship Project for Digital Finance in Europe*, https://www.infinitech-h2020.eu/ (accessed on 1 March 2021). [104]

IOSCO (2020), *The use of artificial intelligence and machine learning by market intermediaries and asset managers Consultation Report INTERNATIONAL ORGANIZATION OF SECURITIES COMMISSIONS*, http://www.iosco.org (accessed on 5 January 2021). [78]

JPMorgan (2019), *Machine Learning in FX*, https://www.jpmorgan.com/solutions/cib/markets/machine-learning-fx (accessed on 14 December 2020). [30]

Kaal, W. (2019), *Financial Technology and Hedge Funds*, https://papers.ssrn.com/abstract=3409548 (accessed on 11 December 2020). [19]

Kalamara, E. et al. (2020), *Making text count: economic forecasting using newspaper text | Bank of England*, https://www.bankofengland.co.uk/working-paper/2020/making-text-count-economic-forecasting-using-newspaper-text (accessed on 15 March 2021). [22]

Klein, A. (2020), *Reducing bias in AI-based financial services*, Brookings, https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/ (accessed on 15 March 2021). [73]

Könberg, M. and M. Lindberg (2001), "Hedge Funds: A Review of Historical Performance", *The Journal of Alternative Investments*, Vol. 4/1, pp. 21-31, http://dx.doi.org/10.3905/jai.2001.318999. [116]

Krizhevsky, A., I. Sutskever and G. Hinton (2017), "ImageNet Classification with Deep Convolutional Neural Networks", *COMMUNICATIONS OF THE ACM*, Vol. 60/6, http://dx.doi.org/10.1145/3065386. [10]

Liew, L. (2020), *What is a Walk-Forward Optimization and How to Run It? - AlgoTrading101 Blog*, Algotrading101, https://algotrading101.com/learn/walk-forward-optimization/ (accessed on 3 December 2020). [26]

Luminovo.ai (2020), *Data Privacy in Machine Learning*, https://luminovo.ai/blog/data-privacy-in-machine-learning (accessed on 30 June 2021). [67]

MAS (2019), *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector Principles to Promote FEAT in the Use of AI and Data Analytics in Singapore's Financial Sector Monetary Authority of Singapore*, https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data (accessed on 15 March 2021). [69]

McKinsey (2020), *Managing and Monitoring Credit risk after COVID-19*, https://www.mckinsey.com/business-functions/risk/our-insights/managing-and-monitoring-credit-risk-after-the-covid-19-pandemic (accessed on 2 September 2020). [113]

Metz, C. (2016), *The Rise of the Artificially Intelligent Hedge Fund | WIRED*, https://www.wired.com/2016/01/the-rise-of-the-artificially-intelligent-hedge-fund/ (accessed on 11 December 2020). [24]

Mollemans, M. (2020), *AI in Sell-Side Equity Algorithms: Survival of the Fittest | TABB Group*, https://research.tabbgroup.com/report/v18-012-ai-sell-side-equity-algorithms-survival-fittest (accessed on 28 June 2021). [28]

New Energy and Industrial Technology Development Organization (2021), *Research on the development of digital technology for regulatory enhancement and Anti-Money Laundering*, https://www.nedo.go.jp/library/database_index.html (accessed on 22 July 2021). [41]

Noonan (1998), *AI in banking: the reality behind the hype | Financial Times*, Financial Times, https://www.ft.com/content/b497a134-2d21-11e8-a34a-7e7563b0b0f4 (accessed on 3 December 2020). [103]

Nourani, M. et al. (2020), *The Effects of Meaningful and Meaningless Explanations on Trust and Perceived System Accuracy in Intelligent Systems*, http://www.aaai.org (accessed on 6 January 2021). [97]

OECD (2020), *Financial Consumer Protection Policy Approaches in the Digital Age*, http://www.oecd.org/finance/Financial-Consumer-Protection-Policy (accessed on 15 December 2020). [117]

OECD (2020), *OECD Business and Finance Outlook 2020: Sustainable and Resilient Finance*, OECD Publishing, Paris, https://dx.doi.org/10.1787/eb61fd29-en. [62]

OECD (2020), *The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector*, https://www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.htm. [6]

OECD (2020), *The Tokenisation of Assets and Potential Implications for Financial Markets*, https://www.oecd.org/finance/The-Tokenisation-of-Assets-and-Potential-Implications-for-Financial-Markets.htm. [53]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://dx.doi.org/10.1787/eedfee77-en. [5]

OECD (2019), *Initial Coin Offerings (ICOs) for SME Financing*, https://www.oecd.org/finance/initial-coin-offerings-for-sme-financing.htm (accessed on 16 September 2020). [55]

OECD (2019), *OECD Business and Finance Outlook 2019: Strengthening Trust in Business*, OECD Publishing, Paris, https://doi.org/10.1787/af784794-en. [37]

OECD (2019), *Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*, https://doi.org/10.1787/d62f618a-en. [4]

OECD (2017), *Algorithms and Collusion: Competition Policy in the Digital Age*, https://www.oecd.org/daf/competition/Algorithms-and-colllusion-competition-policy-in-the-digital-age.pdf. [35]

OECD (2011), *G20 High-Level Principles on Financial Consumer Protection*, https://www.oecd.org/daf/fin/financial-markets/48892010.pdf. [68]

Pearl, J. and E. Bareinboim (2014), "External validity: From do-calculus to transportability across populations", *Statistical Science*, Vol. 29/4, pp. 579-595, http://dx.doi.org/10.1214/14-STS486. [99]

PWC (2020), *Banks' approach to COVID-19: Being data-driven and being human*, https://www.pwc.com/ph/en/business-unusual/banks-approach-to-covid-19.html (accessed on 2 September 2020). [112]

S&P (2019), *Avoiding Garbage in Machine Learning*, https://www.spglobal.com/en/research-insights/featured/avoiding-garbage-in-machine-learning (accessed on 2 September 2020). [48]

Samuel, A. (1959), "Some studies in machine learning using the game of checkers", https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/636026949/report_frank_gabel.pdf (accessed on 11 December 2020). [7]

Scott, S. (2020), *It's not enough for banks to admit misconduct. They've got to prevent it.* | *American Banker*, American Banker, https://www.americanbanker.com/opinion/its-not-enough-for-banks-to-admit-misconduct-theyve-got-to-prevent-it (accessed on 18 March 2021). [13]

The Economist (2019), *Human-machine interface - Data-labelling startups want to help improve corporate AI | Business | The Economist*, https://www.economist.com/business/2019/10/17/data-labelling-startups-want-to-help-improve-corporate-ai (accessed on 16 December 2020). [71]

The Technolawgist (2020), *Does the future of smart contracts depend on artificial intelligence? - The Technolawgist*, https://www.thetechnolawgist.com/2020/12/07/does-the-future-of-smart-contracts-depend-on-artificial-intelligence/ (accessed on 12 January 2021). [56]

UK Information Commissioner's Office (2020), *What are the contextual factors?*, https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/what-are-the-contextual-factors/ (accessed on 15 March 2021). [82]

US Department of Justice (2015), *Futures Trader Charged with Illegally Manipulating Stock Market, Contributing to the May 2010 Market 'Flash Crash' | OPA | Department of Justice*, https://www.justice.gov/opa/pr/futures-trader-charged-illegally-manipulating-stock-market-contributing-may-2010-market-flash (accessed on 18 March 2021). [38]

US Treasury (2018), *A Financial System That Creates Economic Opportunities Nonbank Financials, Fintech, and Innovation Report to President Donald J. Trump Executive Order 13772 on Core Principles for Regulating the United States Financial System Counselor to the Secretary*. [8]

US Treasury (2016), *Opportunities and Challenges in Online Marketplace Lending 10 May 2016*. [45]

Ushida, R. and J. Angel (2021), *Regulatory Considerations on Centralized Aspects of DeFi managed by DAOs*. [60]

Vellido, A., J. Martín-Guerrero and P. Lisboa (2012), *Making machine learning models interpretable*, http://www.i6doc.com/en/livre/?GCOI=28001100967420. (accessed on 11 March 2021). [83]

Wang, S. (2003), "Artificial Neural Network", in *Interdisciplinary Computing in Java Programming*, Springer US, Boston, MA, http://dx.doi.org/10.1007/978-1-4615-0377-4_5. [118]

Weber, J. (2019), *Welcome to the New World of Equity Trade Execution: MiFID II, Algo Wheels and AI | Coalition Greenwich*, https://www.greenwich.com/node/110066 (accessed on 28 June 2021). [29]

Westerhuis, J. et al. (2008), "Assessment of PLSDA cross validation", *Metabolomics*, Vol. 4/1, pp. 81-89, http://dx.doi.org/10.1007/s11306-007-0099-6. [91]

White & Case (2017), *Algorithms and bias: What lenders need to know*, https://www.jdsupra.com/legalnews/algorithms-and-bias-what-lenders-need-67308/ (accessed on 2 September 2020). [46]

Widmer, G. (1996), *Learning in the Presence of Concept Drift and Hidden Contexts*. [94]

World Bank (2020), *The World Bank Group COVID-19 Policy Responses: Why Credit Reporting Matters in the Stabilization and Recovery Phases?*, https://financialit.net/news/covid-19/impact-covid-19-personal-loan-market (accessed on 2 September 2020). [111]

Xu, Y. and R. Goodacre (2018), "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning", *Journal of Analysis and Testing*, Vol. 2/3, pp. 249-262, http://dx.doi.org/10.1007/s41664-018-0068-2. [90]

Zhang, J. and E. Bareinboim (2020), *Can Humans be out of the Loop?*. [100]

Ziqi Chen et al. (2020), *Cortex Blockchain Whitepaper*, https://www.cortexlabs.ai/cortex-blockchain (accessed on 12 January 2021). [54]

Notes

[1] Reinforcement learning AI involves the learning of the algorithm through interaction and feedback. It is based on neural networks and may be applied to unstructured data like images or voice.

[2] Concept drifts describe situations where the statistical properties of the target variable studied by the model change, which changes the very concept of what the model is trying to predict. Data drifts occur when statistical properties of the input data change, affecting the model's predictive power.

[3] Big dataor alternative data or smart data.

[4] The use of the term AI in this note includes AI and its applications through ML models and the use of big data.

[5] Inspired by the functionality of human brains where hundreds of billions of interconnected neurons process information in parallel, neural networks are composed of basic units somewhat analogous to human neurons, with units linked to each other by connections whose strength is modifiable as a result of a learning process or algorithm (Wang, 2003[118]).

[6] For the purposes of this section, asset managers include traditional and alternative asset managers (hedge funds).

[7] It should also be noted that, as many of the new datasets used do not span very long in terms of history, it is difficult to apply empirical statistical analysis.

[8] Algo wheels take the trader bias out of which broker's algorithm gets deployed in the marketplace.

[9] Walk forward optimisation is a process for testing a trading strategy by finding its optimal trading parameters in a certain time period (called the in-sample or training data) and checking the performance of those parameters in the following time period (called the out-of-sample or testing data) (Liew, 2020[26]).

[10] Such tools can also be used in HFT to the extent that investors use them to place trades ahead of competition.

[11] As opposed to value-based trade, which focuses on fundamentals.

[12] Such automated mechanisms are known as 'kill switches' or 'dead man's handles' by the industry.

[13] It should be noted, however, that the risk of discrimination and unfair bias exists equally in traditional, manual credit rating mechanisms, where the human parameter could allow for conscious or unconscious biases.

[14] Oracles feed external data into the blockchain. They can be external service providers in the form of an API end-point, or actual nodes of the chain. They respond to queries of the network with specific data points that they bring from sources external to the network.

[15] Endo (2019), Introductory remarks by Commissioner Toshihide Endo (unpublished).

[16] Reinforcement learning involves the learning of the algorithm through interaction and feedback. It is based on neural networks and may be applied to unstructured data like images or voice.

[17] Contrary to explicit collusion, where the anti-competitive conduct is maintained through explicit agreements, tacit collusion refers to forms of anti-competitive co-ordination achieved without any explicit agreement, but which competitors are able to maintain by recognising their mutual interdependence.

[18] This typically occurs in transparent markets with few market players, where firms can benefit from their collective market power without entering in any explicit communication.

[19] In cases of credit decisions, this also includes information on factors, including personal data that have influenced the applicant's credit scoring. In certain jurisdictions, such as Poland, information should also be provided to the applicant on measures that the applicant can take to improve their creditworthiness.

[20] In the future, the explainability rules for the internal models using AI/ML should be applied for the calculation of both market risk and CCP margins.

www.oecd.org/finance