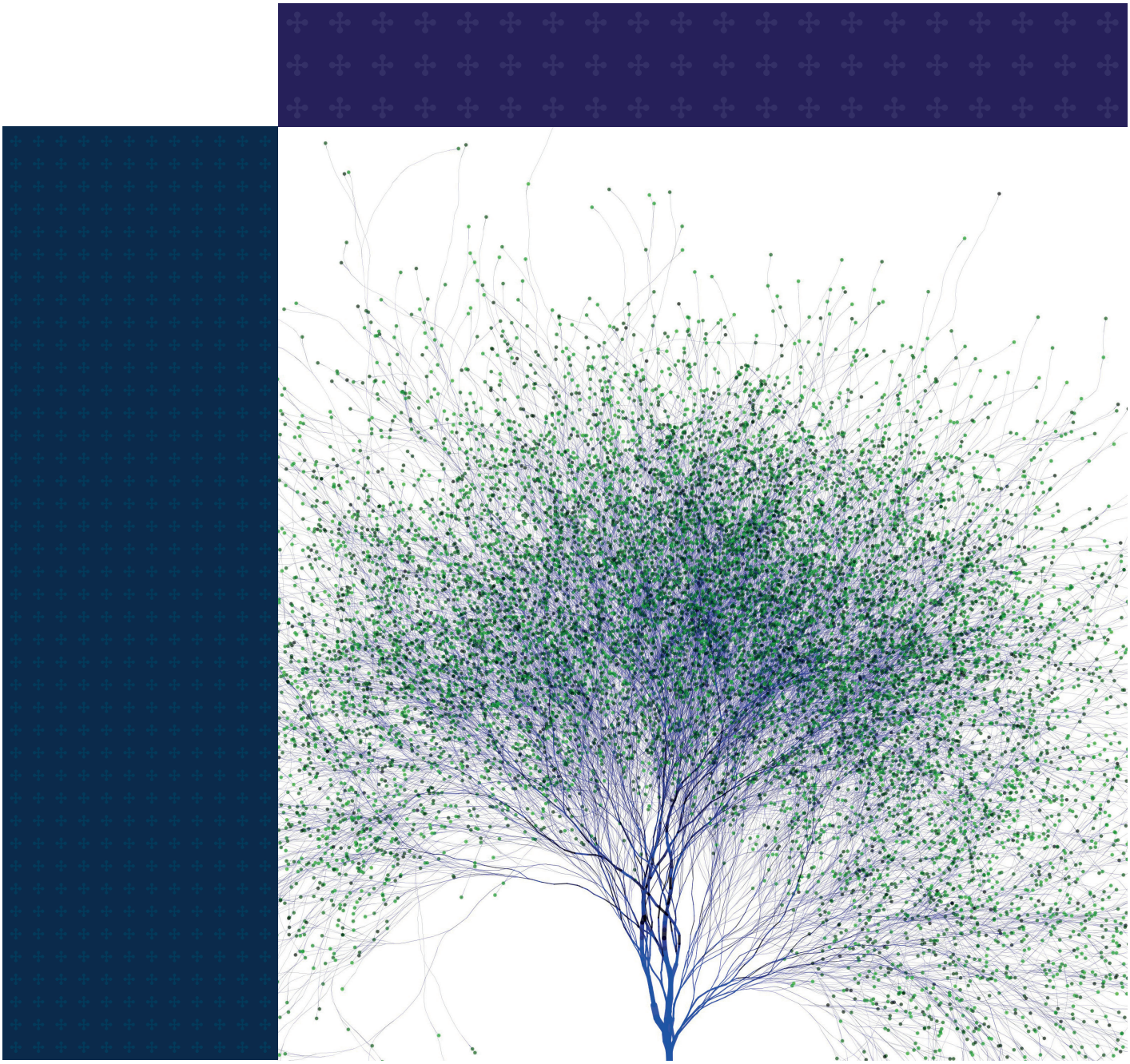
Final report

Artificial Intelligence Public-Private Forum

February 2022



Contents

List of figures

List of boxes

BANK OF ENGLAND

FCA FINANCIAL CONDUCT AUTHORITY

Foreword

Artificial intelligence (AI) is an increasingly important technology for UK financial services. Firms already use AI across a wide range of business activities and the Covid-19 pandemic has accelerated the pace of adoption. The improved classification and predictive accuracy of AI models, as well as their ability to automate certain tasks, can bring benefits for households, firms, and the economy. For example, consumers can potentially access lower-cost and more-tailored financial products and services. It is important that the financial services sector is able to harness the value of AI for the benefit of society at large.

At the same time, innovation can change the trade-offs between risk and return. The speed, scale, and complexity of AI systems can amplify existing risks and pose new challenges. For example, the complexity of AI models can result in greater predictive accuracy but can also make it harder to explain outputs. They can also make it harder to understand how multiple models will interact, especially as they become more interconnected within financial markets.

The benefits and challenges of AI can arise at different points in a product or service lifecycle. For example, when considering the various stages of a loan application, AI systems may be used for marketing the loan, processing the application, guiding or making the loan decision, automating credit stewardship processes, and modelling regulatory capital requirements. While AI serves to make the process quicker and the modelling more accurate, it does raise many challenges. These include potential bias in the data that may exclude certain demographics; lack of transparency and explainability that make it difficult to understand how the AI model's inputs lead to its outputs; and unclear accountability and responsibility for decisions based on those outputs.

Another example is the use of AI in savings and investment management. And as it becomes more widely used, in, for example, institutional fund products, its potential impact on markets may increase. This could lead to 'herd' behaviour if trading systems use similar data and models, or through concentrations and potential disruptions in the networks used to transfer data and models. All of which can potentially pose risks to consumers, firms, and the financial system. These are only two examples of the ways in which AI systems are increasingly being used across the financial system.

To minimise risks and unintended consequences, consumers, firms and regulators need a solid understanding of the technology. That is why the Bank of England (Bank) and the Financial Conduct Authority (FCA) established the AI Public-Private Forum (AIPPF) in October 2020 to further the dialogue between the public sector, the private sector, and academia on AI. The aim of the AIPPF was to share information, deepen our collective understanding of the technology and explore how we can support the safe adoption of AI in financial services.

The issues and topics covered by the AIPPF fell into three areas: Data, Model risk, and Governance. Our previous work has shown that the drivers of AI risk in financial services can occur at each level within interconnected AI systems: starting with the risks associated with use of data to train and run AI models; building into risks arising from the use of AI models themselves; which in turn feed into risks to the firm and the governance structures it needs to manage those risks; and finally, the risks to the financial system as a whole.

This report represents the culmination of more than a year's worth of meetings, workshops, and discussions. All of this work was done entirely virtually, again reflecting the importance of technology to our society, financial sector, and economy. We are immensely grateful to the members and observers for contributing their time, expertise, and knowledge to the AIPPF. This project and report would not have been possible without them.

The purpose of this report is to summarise the key issues discussed by the AIPPF members. In particular, the challenges and risks associated with the use of AI in financial services, including examples of how to address and mitigate them. The report also highlights considerations for the regulatory framework and how it can support the adoption of AI. Given this is an evolving area and there are still many unanswered questions, what "good" looks like can differ depending on the situation.

One thing the AIPPF has made clear to us is that the private sector wants regulators to have a role in supporting the safe adoption of AI in UK financial services, building on what is already in place. Different types and sizes of firms will have different views and needs, and any regulatory interventions should be proportionate. Much work is already underway as regulators, domestically and internationally, consider issues relevant to their respective remits. One of the key questions is how existing regulation and legislation, such as the Senior Managers and Certification Regime, may be applied to AI and whether AI can be managed through extensions of the existing regulatory framework, or whether a new approach is needed.

To help address this question and support further discussion about what an appropriate role for regulators might look like, we will publish a Discussion Paper on AI later this year. It will build on the work of the AIPPF and broaden our engagement to a wider set of stakeholders. Discussion Papers are used to stimulate debate on issues about which we are considering making rules or setting out expectations. The Discussion Paper will aim to provide clarity around the current regulatory framework and how it applies to AI, ask questions about how policy can best support further safe AI adoption, and give stakeholders an opportunity to share their views. The responses to the Discussion Paper will help us to identify what is most relevant to our remits and what is not, as well as help formulate any potential policy. In the meantime, we will continue to work with the Government and other regulators in the UK and elsewhere to support the safe adoption of AI.

The AIPPF has provided a very helpful contribution to the broader debate on AI. We hope that this report will advance the collective understanding of how AI is used in UK financial services and serve as the foundation for future collaboration with academics, consumers, firms, and regulators around the world.

Dave Ramsden
Deputy Governor, Markets and Banking
Bank of England

BANK OF ENGLAND

Jessica Rusu
Chief Data, Information and Intelligence Officer
Financial Conduct Authority

FCA FINANCIAL CONDUCT AUTHORITY

Establishing the AI Public-Private Forum

AI Public-Private Forum

1. Artificial intelligence (AI) is increasingly used in UK financial services and Covid-19 has accelerated the pace of adoption.(1) The technology can bring a range of benefits to consumers, firms, and the wider financial system; but there are also barriers to adoption and challenges. The use of AI in financial services may also create new risks or amplify existing ones. Therefore, financial services firms need to keep up with appropriate controls and focus on the resilience of their AI systems. At the same time, clarity from regulators about their expectations is a critical part of fostering innovation and may support safe adoption. Engagement between firms and regulators is one way to address these issues.

2. That is why the Bank of England in its response to the Future of Finance Review(2) announced in July 2019 (3) that the Bank and FCA would establish the AI Public-Private Forum (AIPPF) to further dialogue on AI innovation and its safe adoption within financial services. More specifically the AIPPF sought to:

2.1. Share information and understand the practical challenges of using AI within financial services, as well as the barriers to deployment and potential risks.

2.2. Gather views on potential areas where principles, guidance or good practice examples could be useful in supporting safe adoption of these technologies.

2.3. Consider whether ongoing industry input could be useful and what form this could take.

3. In January 2020, the Bank and FCA published the AIPPF Terms of Reference, which included a call for Expressions of Interest.(4) The Bank and FCA received more than 100 applications from a wide range of AI experts. After reviewing the applications against the selection criteria and seeking input from an independent senior adviser to the Prudential Regulation Authority (PRA), the Bank and FCA selected 21 members.(5)

4. The diverse group of members were drawn from a range of sectors. They have extensive knowledge of this complex field and relevant backgrounds, having worked with AI. Members acted in a personal capacity, with a view to furthering the interests of the market as a whole, rather than representing the views or interests of their employers or organisations.

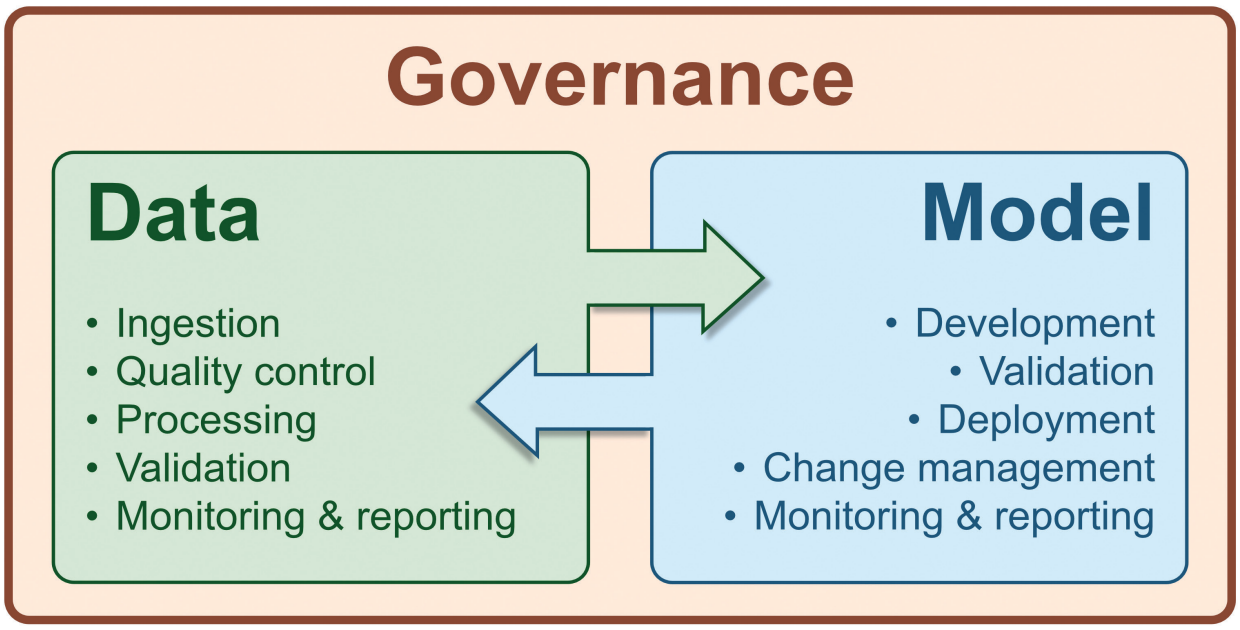
5. In addition to the private sector and academic members, five observer organisations were invited to join the AIPPF. The observers included the Centre for Data Ethics and Innovation (CDEI), the Fixed Income, Currencies and Commodities Markets Standards Board (FMSB), HM Treasury (HMT), the Information Commissioner's Office (ICO), and the Office for Artificial Intelligence (OAI).

6. The AIPPF was co-chaired by Dave Ramsden, Deputy Governor for Markets and Banking at the Bank, and initially by Sheldon Mills, Executive Director for Consumers and Competition, then Jessica Rusu, Chief Data, Information and Intelligence Officer, at the FCA. The AIPPF was co-ordinated and facilitated by a secretariat drawn from staff of the Bank and FCA.(6)

(1) The impact of Covid on machine learning and data science in UK banking (Bank of England); AI adoption accelerated during the pandemic (KPMG).
(2) Facilitating firms' use of technology, like the cloud, to increase their operational resilience (Bank of England).
(3) Enable, empower, ensure: a new finance for the new economy, speech by Mark Carney (Bank of England).
(4) Artificial Intelligence Public-Private Forum: Terms of Reference (Bank of England).
(5) AIPPF, List of Members.
(6) The AIPPF secretariat consisted of Kathleen Blake, Mohammed Gharbawi, Varun Paul, Oliver Thew, and Seema Visavadia from the Bank, and Leo Gosland and Henrike Mueller from the FCA.

7. The AIPPF launched on 12 October 2020 (7) and ran for one year, with four quarterly meetings and a number of workshops structured around the three key topics of: Data, Model risk, and Governance. This structure was used because the drivers of AI risk in financial services, and many of the challenges that firms face, can occur at each of these levels within interconnected AI systems (see Figure 1). Essentially, issues at the data level can impact the models, which then raise broader governance challenges. The same structure is used for this report, with key chapters on Data, Model risk, and Governance.

Figure 1: AI system



Sources: AIPPF.

Purpose of this report

8. As stated in the Terms of Reference, the AIPPF committed to deliver a public document on its conclusion. This report meets that commitment and also aims to advance the collective understanding of the use of AI in UK financial services, as well as promote further debate among academics, practitioners, and regulators about how best to support safe adoption of this technology.

9. Please note that this report is not intended to be a primer or comprehensive overview on the use of AI in financial services, as there are other relevant publications.(8) Instead, it presents the culmination of a year-long dialogue between the AIPPF members, observers, and secretariat.

10. More specifically, this report explores the various barriers to adoption, challenges and risks related to the use of AI in financial services. In order to forward the debate and understanding of AI, it also explores potential ways to address such barriers and challenges, as well as mitigate potential risks. Given that this is a rapidly evolving area, there are no clear answers yet. As the AIPPF has a wide range of expertise, the report presents various ways to address challenges and suggests different approaches to mitigate risks. The use of AI in financial services also has implications for a number of regulators, rather than just the Bank and FCA. For example, the Digital Regulation Cooperation Forum brings together a number of UK regulators to ensure greater cooperation on online regulatory matters, such as AI.(9)

(7) Fintech AI Public-Private Forum (Bank of England).
(8) Machine learning in UK financial services (Bank of England); Artificial intelligence and machine learning in financial services (Financial Stability Board); AI in Business and Finance (OECD); AI in financial services (Alan Turing Institute); Final Report on Big Data and Advanced Analytics (European Banking Authority).
(9) The Digital Regulation Cooperation Forum (GOV.UK).

Disclaimer

This report and all content apart from the Foreword and the 'Establishing the AI Public-Private Forum' section reflect the views of the AIPPF members. The views expressed by the AIPPF members in this report do not reflect the views of their institutions, since the members acted in a personal capacity and contributed to the AIPPF with the aim of furthering the interests of the market as a whole, rather than representing the views or interests of their individual organisations.

This report does not reflect the views of the Bank or the FCA. Neither the Bank nor FCA, or any of their staff, officials, or representatives, including but not limited to the AIPPF co-chairs, are responsible for any of the views or statements expressed in this report.

Moreover, this report and the activities, discussions, and outputs of the AIPPF should not be taken as an indication of future policy by the Bank or FCA. Bank and FCA policy positions will continue to be developed and communicated in accordance with the usual governance and public consultation procedures of the two organisations.

Executive summary

1. **Artificial intelligence (AI) is a rapidly evolving and powerful tool which financial services firms are using in an increasing number of ways**. It can bring benefits to consumers, businesses, and the wider economy, but AI can also amplify risks and create new challenges. The AI models used in the financial system are becoming increasing sophisticated. Their speed, scale, and complexity, as well as their capacity for autonomous decision-making, have already sparked considerable debate.

2. **The Bank of England (Bank) and the Financial Conduct Authority (FCA) established the AI Public-Private Forum (AIPPF) in 2020**[10] **to further the discussion between the public sector, the private sector, and academia, and to encourage the safe adoption of AI in financial services**. The AIPPF discussions explored the benefits and the risks of AI, and ways in which the public and private sectors can support the safe adoption and use of AI in UK financial services. Those discussions started with data, as the foundation of AI, then moved into models, where the data are put to work, and ended with governance, which provides the guardrails.

3. This executive summary highlights the key findings under each of those headings. The introduction explains the background to the AIPPF, with subsequent chapters reflecting the AIPPF discussions and providing detailed insights and best practice on Data, Model risk, and Governance in turn.

Data

4. **AI begins with data**. The recent growth in AI use has been partly driven by the vast increase in the data available to train AI models with. While modelling choices and the management of model risk are clear priorities, many of the benefits and risks can be traced back to the data, rather than the AI systems or algorithms themselves.

5. **Data quality is as important as ever**. At such scale, there is a premium on understanding the attributes of the data being deployed, including provenance, completeness, and how representative they are. Also, because the way data are used can change, the importance of documentation, versioning, and ongoing monitoring is increasing.

6. **One of the defining features of AI is its ability to process large volumes of unstructured or 'alternative' data**. These may come from different sources, including satellite images, biometrics, or telematics, which can generate unique insights. But this use of alternative data is also one of the main ways that AI could increase data quality issues. This is partly because these data are often sourced from third-party providers, presenting additional challenges of quality, provenance, and sometimes, legality.

7. **Data are also reshaping organisations**. The changing role of data in the AI lifecycle raises questions on the appropriate governance structures within an organisation. And as firms develop their data strategies to accommodate the use of AI systems, there is an increasing call for the development and use of AI-specific data standards.

Model risk

8. **Most of the risks related to the use of AI models in financial services are not new** and can arise from the use of non-AI models. It is the scale at which AI is beginning to be used, the speed at which AI systems operate, and the complexity of the underlying models which is new. These too can create new challenges as well as amplify existing ones.

9. **Complexity is the key challenge for managing the risks arising from AI models**. This includes the complexity of the inputs (such as many input layers and dimensions), complex relationships between variables, the intricacies of the

(10) Fintech AI Public-Private Forum (Bank of England).

models themselves (e.g. deep learning models), and the types of outputs, which may be actions, algorithms, unstructured (e.g. images or text), and/or quantitative. It becomes even more complicated when several AI models operate within a network.

10. **Being able to explain model outputs is vital**. Approaches to managing explainability should not just focus on the features and parameters of models, but also on consumer engagement and clear communications.

11. **Identifying and managing change in AI models, as well as monitoring and reporting their performance, are keys parts of ensuring that models behave as expected**. This includes monitoring for changes in a model's functional form as well as its outputs and potential need for retraining.

Governance

12. **A key characteristic of AI systems is their capacity for autonomous decision-making**. This can have profound implications for how to govern the technology and its outcomes, including ensuring effective accountability and responsibility.

13. **Existing governance frameworks and structures provide a good starting point for AI models and systems**, partly because AI models will invariably interact with other risk and governance processes. The most relevant of these in financial services are data governance and model risk management (MRM) frameworks, as well as operational risk management.

14. **Governance of AI systems should reflect the risk and materiality of the use-case** (i.e. the specific application for which the AI systems are designed), even when existing governance frameworks are applied to AI. Where possible, firms should leverage and adapt existing governance structure to manage AI, including data and MRM frameworks.

15. **A centralised body within firms should set the AI governance standards**. Overall responsibility for AI could be held by one or more senior managers, with business areas being accountable for the outputs, compliance, and execution against the governance standards. Transparency and communication are key elements of AI governance.

16. **Governance of AI in firms is more effective if it includes diversity of skills and perspective, and if it covers the full range of functions and business units.** This type of cross-functional approach can help manage the complexity of AI systems and their associated data challenges.

17. **Firms should ensure that there is an appropriate level of understanding** and awareness of AI's benefits and risks throughout the organisation. This can include more formal governance arrangements and a clear role for senior leadership.

Next steps

18. **The discussion on the safe adoption of AI has only just begun**. The open dialogue between the public and private sectors through the AIPPF has been informative and productive. There is considerable appetite for this dialogue to continue in some form beyond the AIPPF to support the safe adoption of AI in financial services.

19. **An industry body for practitioners could build trust in the use of AI**. Professionalising data science, including voluntary codes of conduct and establishing an auditing regime would help foster wider acceptance of and trust in AI systems.

20. **Regulators can support innovation and adoption of AI**. Regulators could start by providing clarity on how existing regulation and policies apply to AI. To support innovation, guidance should focus on the outcomes expected and not be overly prescriptive. It could provide illustrative case studies and identify high-risk use-cases to focus on.

21. **Regulatory alignment will catalyse progress**. The coordination between the Bank and FCA has been very helpful, and continued coordination with domestic regulators and government departments will be crucial. But transnational coordination between regulators is also needed to keep all parts of the global AI ecosystem in check.

Introduction

Definition and characteristics of artificial intelligence

22. AI is difficult to define. However, it is important to give a sense of what customers, firms, and regulators mean when they use the term. Providing clarity can help demystify AI and separate the hype from reality. This can, in turn also (i) help customers understand how and when they are being affected by the use of AI, (ii) enable firms to use the technology responsibly and assess the relevant trade-offs, (iii) help with the practical implementation of high-level principles, and (iv) lead to more standardised approaches across industry. All of which can support wider and safer adoption of AI in financial services.

23. In the broadest sense, AI is the use of advanced statistical techniques with large computational and data needs. More specifically, AI is often used to describe machine learning (ML). ML is a sub-set of AI that can be broken down into three main categories: supervised, unsupervised, and reinforcement learning.(11) The definition provided in the draft European Commission (EC) regulation on AI (12) is broad and includes statistical models and techniques that are not always considered to be AI. The Bank and FCA previously defined ML as the theory and development of computer systems able to perform tasks which require human intelligence.(13)

Box 1: Definition of AI used in this report

This report uses the International Organization for Standardization (ISO) definition of AI,(14) which is "an interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning." In this report the term AI also refers to ML.

24. As well as defining AI, it is important to consider the characteristics of AI applications and how they differ from non-AI applications that produce the same result. These characteristics may include the complexity of AI, its iterative approach, the use of hyperparameters,(15) and the use of unstructured datasets. The Bundesbank and BaFin, for example, used this approach in a recent publication (see Figure 2).(16) Rather than use a specific definition of ML, the paper sets out various ML characteristics that create a boundary of what could be considered within and out of scope. In the absence of any single recognised definition of AI, this may be a helpful approach for firms and regulators.

(11) Supervised, Unsupervised, & Reinforcement Learning (AI Wiki).
(12) EC proposal for regulation of AI (EUR-Lex - 52021PC0206).
(13) Machine learning in UK financial services (Bank of England).
(14) ISO/IEC 2382:2015(en), Information technology - Vocabulary (ISO).
(15) Understanding Hyperparameters and Optimisation techniques (Towards Data Science).
(16) Consultation Paper: Machine learning in risk models - Characteristics and supervisory priorities (Bundesbank).

Figure 2: Machine Learning characteristics

Sources: Bundesbank and BaFin.

Box 2: Hyperparameters

Hyperparameters are parameters used in AI to control the model selection and training process. They are specified before training iterations, unlike parameter values which are calculated as part of the model fitting process.

Artificial intelligence in financial services

25. Although this report does not provide a comprehensive overview of the use of AI in financial services, it is useful to have a general understanding of the technology, its application in financial services and the benefits and risks (see Annex 1 for specific use-cases of AI in financial services). This will help put the following chapters on Data, Model risk, and Governance in context.

26. Data are at the core of financial services and the industry has always used mathematical and statistical modelling to provide insights, support, and deliver financial decisions. So it is unsurprising that finance is one of the sectors that has seen relatively widespread adoption of AI. From customer services to consumer credit, anti-money laundering (AML) and anti-fraud analytics to investment management, financial services firms use AI for a range of business services.(17)

27. As with any technology, AI maturity and adoption varies depending on the firm and sector. However, Covid-19 has accelerated the overall pace of AI adoption both for in-house models and third-party providers,(18) as well as the wider shift towards an online society and economy. AI platforms and open-source libraries are enabling even greater access to AI algorithms, with data being the key differentiator. Similarly, the technology is developing at a rapid pace. Decentralised and networked systems are playing a greater part in AI systems, with the role of federated learning(19) and 'bringing computation to data' increasing.

(17) Machine learning in UK financial services (Bank of England); AI Barometer (CDEI).
(18) The impact of Covid on machine learning and data science in UK banking (Bank of England); AI adoption accelerated during the pandemic (KPMG).
(19) A method of training models across distributed and/or decentralised networks using local data and without exchanging those data.

28. In most cases, firms use AI to augment or upgrade from existing rules-based models. In some cases, the AI techniques (e.g. tree-based decision models) are only slightly more complex than the existing rules-based techniques (e.g. gamma distribution). Moreover, many of the issues covered in this report are not specific to AI (such as data governance or model risk) and so need to be seen with a wider lens than just AI.

29. In many ways, the use of AI in financial services can be seen as a continuation of existing modelling. For example, most AI applications do not replace the full scope of the traditional process but rather augment existing analytical techniques. Also, in most cases there is still a human-in-the-loop at some stage.(20) In other words, we are not yet seeing full-scale automation of decision-making in UK financial services. Although there is a continuum of complexity, it is worth considering '*what changes with AI*'?

Benefits for consumers, firms, and the financial system

30. AI can bring significant benefits to consumers, firms, and the wider financial system.(21) For customers, AI can enable more personalised financial products and services, including savings and investment advice. AI can also enable a more seamless customer journey and experience, with the use of Natural Language Processing (NLP), as well as voice, document, image, and facial recognition.

31. For firms, AI can help improve predictive power and increase profitability. Firms are already using AI to improve operational efficiency, which helps to reduce processing time and reduce costs as clients demand faster and more streamlined and robust services. From monitoring risks to regulatory reporting, a range of front and back office operations can be automated. Firms can focus on areas where repetitive and manual tasks can be automated, freeing up subject matter experts to work on more complex tasks. The combination of AI and real-time data also has the potential to lead to more effective decisions.

32. The benefits for firms can also extend to the financial system and wider economy. For example, AI can help analyse and make sense of the scale and complexity of the financial system, especially in cases where individuals and firms are managing levels of complexity beyond human intellect. AI can also help tackle synthetic identity fraud, where synthetic identities are created from a jigsaw of real data, by identifying cases which may be difficult to identify by human analysts.

Barriers to adoption, challenges, and risks

33. Clearly, there are many business areas and use-cases where firms can use AI. However, there can be significant barriers to choosing and using AI systems, especially for smaller organisations with fewer resources. Some of the biggest barriers involve data, including access to data and managing data privacy. At the model level, there are challenges in explaining and documenting the workings and outputs of complex models, as well as ensuring appropriate governance around using such data and models. More broadly, regulatory uncertainty is another significant barrier.

34. A further barrier can be firms' ability to identify appropriate use-cases for AI, especially when first starting to explore and use the technology. It may be more useful to look at small-scale use-cases where the firm already has enough data and the necessary buy-in from the business. This may be considered good practice for firms.

35. As with any technology, using AI in financial services can pose risks to consumers, firms, the financial system, and the wider economy.(22) Although risk management in financial services has improved over time, AI can amplify existing risks and introduce new challenges that mean existing controls and frameworks may need to evolve to effectively manage the risks of using AI. For example, bias embedded in historical data may cause algorithms to give more importance to certain features, which then make incorrect predictions.

36. The use of ever larger data sets that are harder to validate at a granular level could, for example, lead to unexpected outcomes and risks, like biased and unfair decisions. There are also challenges in using multiple internal

(20) Artificial Intelligence in Finance: Putting the Human in the Loop (Zetzsche, et al. 2020).
(21) See a mapping of benefits and risks in AI in Financial Services (Alan Turing Institute).
(22) See the taxonomy and hierarchy of risks in Artificial Intelligence Risk and Governance (ARIS Group).

data sets that were previously kept in silos, as well as third-party datasets, that may require updated data management. Some AI models are also dynamic, meaning they continuously learn from live data and so their outputs can change. This is a particular issue when the underlying data inputs change (data drift) or statistical properties of the data change (concept drift), which can affect all types of models and create different risks. Equally, model performance can deteriorate due to incorrect training data. So more continuous monitoring and validation (and potentially documentation) may be required to manage this risk, compared to static validation and testing methods.

37. AI can lead to more tailored financial services and products. However, there are trade-offs with this. For example, highly tailored products and services can lead to some consumers being priced out. Customers may also be unable to get credit or insurance cover if they are deemed to be higher risk. Other risks to consumers include competition and fiduciary concerns, such as unfavourable and/or unfair penalties or product conditions (such as level of collateral needed). The combination of AI and personal data could also risk breaching consumers' personal data rights.

38. For firms, inappropriate use of AI could result in financial loss (e.g. from a poor credit algorithm), reputational risk, regulatory risk (e.g. being fined or sanctioned because of breach of responsibilities to customer), operational risk (e.g. being unable to recover or debug), and risk of loss of intellectual property. Certain adaptive AI models need large amounts of data and are open to 'adversarial attacks',(23) which can expose firms to even greater cyber risks. Another challenge for firms is that AI can introduce non-financial risks that are less well understood, such as those involving data privacy and protection. This is important to consider as protected characteristics may be needed to measure or establish the fairness of AI predictions.

Domestic and international context

39. AI will likely become an increasingly integral technology for various segments of the UK economy and wider society. This is recognised by the UK Government in the UK National AI Strategy, which was published by the OAI in September 2021.(24) The National AI Strategy(25) builds on several other key initiatives, such as the National Data Strategy, and aims to establish a ten-year plan to make the UK an 'AI superpower'. The CDEI has also published leading research on a number of relevant areas, like bias in algorithmic decision-making,(26) to help promote safe and ethical adoption of AI in the UK in recent years. Similarly, the FMSB has published papers on the emerging themes and challenges in algorithmic trading, to help market participants address some of the potential AI-related challenges.(27)

40. Some UK regulators have tried to address the use of AI in their respective remits. In particular, the ICO has issued guidance on explainability that can help firms use AI in a way that complies with data protection laws.(28) The guidance outlines different levels of explainability, depending on the domain, the use-case and the impact on the individual. Ultimately, the aim is to provide reassurance around the governance and provenance and use of personal data, and the link between input and output, without having to delve into the black box of complex AI models (in so far as this is possible).

41. A significant number of AI governance principles have been produced around the world, both for financial services and other sectors.(29) There is broad consensus on the value of these principles, most of which address a similar set of topics. However, the difficulty for firms is in applying them to specific AI use-cases and translating them into effective internal practices. Further regulatory clarification on how best to apply high-level principles to specific use-cases would be welcomed by firms. Any guidance for financial services could also seek to include relevant standards and guidance produced in other sectors (such as the ISO and the International Electrotechnical Commission (IEC) standards on AI(30) or the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA) guide for federated learning).(31)

(23) Attacking machine learning with adversarial examples (OpenAI).
(24) National AI Strategy (GOV.UK).
(25) National Data Strategy (GOV.UK).
(26) CDEI review into bias in algorithmic decision-making (CDEI).
(27) Emerging themes and challenges in algorithmic-trading and machine learning (FSMB).
(28) Explaining decisions made with AI (ICO).
(29) AI Ethics Guidelines Global Inventory (Algorithm Watch).
(30) ISO/IEC 2382-28:1995, Information technology – Part 28: Artificial intelligence – Basic concepts and expert systems (ISO).
(31) IEEE 3652.1-2020 (IEEE SA).

42. Other jurisdictions have also progressed their approach to regulating the use of AI in financial services. In particular, the EC proposal for draft regulation on AI was published in April 2021.[32] The documentation requirements of the proposed EC regulation and, in contrast to the wide definition of AI, the narrow definition of high-risk use-cases may be useful for other jurisdictions to consider.

43. Given these developments, there may be a need to avoid regulatory fragmentation where possible, both domestically and internationally, and between different sectors. Harmonising regulation at these levels would help ensure accountability and manage risks without stifling innovation. Especially as contradictions and vagueness make it harder to understand where the differences are between foreign and UK regulations and impede the use of AI models developed outside of the UK.

44. At the same time, the topic of possible regulatory responses to AI is complex. There is a risk that regulation will be too strict and too early. Instead, any regulation should aim to be flexible and it may be that principles-based is the most effective form. Moreover, many current standards and regulations not related to AI may be suitable for the use of AI, and may only require tweaks or clarification. These existing regulations are often unrelated and there is no specific, single source that AI practitioners and firms can use for specific guidance.

45. AI for financial services will become increasingly important as the technology advances and becomes more readily available. Whatever form the underlying technology and infrastructure take, the list of applications and use-cases is likely to grow; as are their benefits and potential risks. It is therefore important that regulators continue to monitor, analyse and assess the evolution of AI to understand how best to support its safe adoption while identifying and helping to manage its risks.

(32) EC proposal for regulation of AI (EUR-Lex - 52021PC0206).

Data

46. The second AIPPF meeting[33] and subsequent workshops explored the role of data in AI and the issues they raise, as well as the many practical data-related challenges that can hinder financial services firms adopting and using AI.

47. One of the defining features of AI is its ability to process large volumes of data and to detect and exploit patterns in those data. While many data considerations are not specific to AI, they are of paramount importance to its adoption and use. This is reflected not just in the need for maintaining data quality, but also more broadly in data strategy, management, and governance.

48. There has been a shift among AI practitioners from a model-centric approach to a data-centric approach.[34] Preparing and constructing datasets is often time-consuming and expensive, having a direct impact on the success of AI projects. Data preparation tasks such as cleansing and removing duplication, as well as identifying and fixing errors, can take 45% of data scientists' time on average.[35] Consistency in annotating datasets can have a greater impact on predictive performance than model tuning.

49. From health diagnostics to e-commerce platforms, AI applications often use relatively simple clustering, time-series prediction or regression models. Their success is usually underpinned by high-quality data rather than highly complex algorithms. So high-quality data, along with fair and equal access to those data, may be a greater enabler (in terms of competition and innovation) than the AI algorithms and models themselves.

Key findings – Data

AI begins with data. One of the defining characteristics of AI is its ability to process large volumes of data and exploit patters that are often hidden. While modelling choices and the management of model risk are clear priorities, many of the benefits and risks from AI can be traced back to the data rather than the AI models, algorithms, and systems.

AI is often used to process large volumes of unstructured or 'alternative' data. These may come from numerous sources such as satellite images, biometrics, or telematics. The use of alternative data by AI systems is one of the main ways that AI could exacerbate data quality issues. This is partly because these data are often sourced from third-party providers, which presents additional challenges relating to quality, provenance, and sometimes, legality.

Data quality processes can raise AI-specific challenges. These are often due to the complexity of data sources and structures used throughout the AI model lifecycle and include the increased importance of documentation and versioning for data and code; ensuring data quality and provenance, completeness and representativeness; and ensuring consistent monitoring.

The value a firm puts on data can influence business models as well as the alignment of its data strategy to its business strategy. Firms should consider pricing of data and expected payoffs from the use of these data in any AI project.

Data ownership and accountability structures across the organisation play a key part in the overall data governance structure. The use of AI raises questions on where to place and manage an appropriate split between data ownership and AI ownership.

(33) Artificial Intelligence Public-Private Forum – Minutes 26 February 2021 (Bank of England).
(34) From model-centric to data-centric AI (Ng, 2021).
(35) 2020 State of Data Science Survey Results (Anaconda).

As firms develop and evolve their data strategies to incorporate the use of AI systems, there is an increasing call for the development and use of data standards specific to an AI context.

Regulatory alignment, in particular transnational coordination between regulators is needed to help standardise solutions offered by data suppliers and/or data marketplaces. International cooperation between data suppliers could also help address issues of fair and equal access to data.

Data quality

Box 3: Data quality

Data quality can be defined in many ways but is generally taken to refer to measures (including accuracy, completeness, consistency, representativeness, and timeliness) of how suitable the data are for a particular purpose.

50. Data quality flows across the AI lifecycle, spanning a range of tasks. These range from profiling data to ensuring metrics are met, through to data preparation, labelling, and quality improvement tasks such as cleansing, removing duplications, to feature engineering (i.e. the process of designing or extracting characteristics or attributes from raw data that are used as model inputs).

51. The complexity of data sources and structures used throughout the AI model lifecycle can present a number of challenges specific to data quality processes for AI systems. For example, additional data quality standards or metrics may be needed (e.g. representativeness), or there may be an increased importance on documentation and versioning for both data and code. A lack of appropriate data science skills may affect data quality and prevent consistent monitoring.

52. A key factor behind many data quality challenges is the attempt to retrofit existing process, controls and systems to AI. Pre-existing ways of managing and governing data quality may not always be appropriate for AI models. This is because existing control frameworks may not scale to the volume and variety of features in the data or its range of applications. As AI systems increase in scale, there is a need for systematic data quality processes, which are transparent, reproducible, auditable, and can be integrated with other processes.

53. Adapting existing data quality metrics and standards to an AI context may also be hampered by a lack of industry-wide consensus on data standards in general, including agreement on good practice. And the rapid pace of change might limit the value of standards-setting organisations, if standards become dated or lose relevance.

54. However, one approach could be for firms to develop their own internal standards and systems to assess data quality. Firms could also develop a template for data 'lineage' to help provide a clear history of where and how data was produced, as well as how it travels through an organisation. This approach may help ensure data are appropriately catalogued and should assist with any data quality impact analysis.

55. The importance of data quality is likely to increase with further data sharing and open-finance initiatives. In effect, achieving a high level of data quality is not cost-free and some approaches to open-data sharing may discourage some firms from investing sufficiently in assuring the quality of their data assets. Data quality can differ significantly between firms and providers in an open-data environment. Coupled with data quality is the need to integrate external data (such as alternative datasets) with internal data. Work being done on data standards as part of Open Banking[36] may be useful for ensuring appropriate data quality, which could support further use of AI in financial services.[37]

(36) Open Banking (OBIE).
(37) FS21/7: Open finance – feedback statement (FCA).

Data strategy and economics

56. The value a firm puts on data can influence its business models, the alignment of its data strategy to business strategy and its use of AI. Firms should consider the pricing of data and expected payoffs from its use in any AI project. The source of the data is a factor not only for questions of consistency and transparency, but also in any cost/benefit analysis.

57. Data can also provide a competitive advantage. While start-ups and new entrants may be able to use new technologies with greater agility, incumbents have a greater volume and variety of data on their side. However, assessing the overall costs (and expected benefits) of individual data assets is not always straightforward. The challenge becomes even greater when using third-party data, especially if data marketplaces are monopolistic or restrictive, or where access to datasets may be limited. This could also create challenges for data pricing policies in individual firms and wider markets.

58. Understanding, measuring and tracking data flows within (as well as into and out of) an organisation is an important part of its data strategy. This includes real-time and streamed data, as well as third-party datasets. Implementing and governing ownership, liability, and control of data in a highly complex distributed system where data are constantly transferred and modified can present many challenges. Data audits and assessments of data usage and alignment with business objectives could form part of the solution.

59. There are also challenges in understanding the provenance and legal status of data sourced from vendors that scrape website data or collate it from a range of sources. This can create risks for firms and may have implications for consumers. These include questions on what data customers are willing to give up for free (e.g. via social media sites or consent to cookies), and how those may be used both in financial services and other sectors.

60. One possible answer to the challenges around data access and differential data pricing is greater use of data marketplaces or exchanges.[38] Such data marketplaces or exchanges could be run solely by one organisation, with the same firm acting as the supplier for all data. Alternatively, the marketplace could operate as an intermediary and allow any party to contribute data. Whatever the model, marketplaces should enable greater access to data. A further concern is the potential unexpected outcomes from the use of data intended for different purposes.

61. So the role of regulation in data access and data pricing must be considered within a broader economic context. For example, there has been a longstanding discussion in the European Union (EU) about the use of consolidated tapes[39] in financial markets and the challenges for international regulation, including distribution of revenue to different participants.

Alternative, unstructured, and synthetic data

62. Perhaps one of the defining features of AI systems is their ability to process large volumes of unstructured or 'alternative' data (see Figure 3). These may come from numerous sources, such as satellite images, biometrics, telematics or shopping patterns, for example. The use of alternative data by AI systems is one of the main ways that AI could amplify data quality issues. In part, this is because these data are often sourced from third-party providers.

(38) A data marketplace is an online site where consumers can view, buy, or sell catalogues of data with associated documentation.
(39) Action 14: Consolidated tape, 2020 action plan (EC).

Figure 3: Data types

Data type	Description	Example
Structured data	• Highly organised • Data objects have fixed meaning • Eg Relational databases or data organised in tabular format	**Standard financial database** First name / Second name / Age / Account balance A / B / 57 / 334 X / Y / 28 / 5,536

Semi-structured data	• Less organised than structured data, some hierarchy (tags, structure) present
	• Some data objects without fixed meaning
	• Eg HTML, JSON, XML
		Website
		`<!DOCTYPE html>`
		`<html>`
		`<head>`
		`<title>Page Title</title>`
		`</head>`
		`<body>`
		`<a href = "URL">Your text / button here"</a>`
		`<button>Your Text Here</button>`

Unstructured data	• Least organised • Information that does not follow a pre-existing data model • Requires analytical techniques to transform it into meaningful information	**Images or text** 

Source: Bank of England.

63. The extent and associated costs of data cleansing needed for effective use of unstructured data can be significant, but this is an essential stage in ensuring that models are trained on accurate and relevant data. Validating aggregated data without knowing the granular structure of those data also presents a risk, as it may contain biases and inaccuracies.

64. Synthetic data, being data generated algorithmically rather than from actual events or direct measurement, as well as aggregated or otherwise processed data, also fall within this area. There is currently a huge amount of interest in generating synthetic data as a way of sharing data while preserving privacy. Synthetic datasets are also being used in building and testing models where there are insufficient 'real' data.

65. It is relatively easy to use synthetic data to train AI models but it may take longer to ensure that these models are suitable for use because models trained on synthetic data also need to be tested on actual data. Setting up suitable pipelines so that synthetic data can be generated and explored quickly with a separate process for validating models on real data is needed to ensure quality, accuracy, and to improve speed to market.

Data governance and privacy

66. Data ownership and accountability structures across the organisation, including data consent management, play a key part in the overall data governance structure. As the chapter on Governance explores in more detail, the use of AI raises questions on where an appropriate split between data ownership and AI ownership should be. For example, should firms create specific data roles, AI data roles, or joint roles? Whatever the split in roles, the organisation should also highlight and ensure collective responsibility.

67. Firms' data management and governance is sometimes organised in silos. Given the increasing reliance on data insights and analytics, this can be inefficient from a governance perspective, particularly with AI systems which need a more holistic, cross-functional approach. Silos and lack of internal access to data also create a risk of inconsistent behaviours and processes.

68. Managing data privacy, data access, and availability is a key element of any data governance framework. So is ensuring full compliance with data protection and additional requirements on using customer data. There may also be additional privacy risks that emerge when individual datasets are combined with other (internal or external) sources, which are not covered by the same decision process.

69. There is also a need to increase data skills across different business areas and teams. Otherwise having a human-in-the-loop might not be an effective safeguard. Similarly, Board members and senior managers are not always aware of, or do not fully appreciate, the importance of issues like data quality. There is a need to increase understanding and awareness at all levels of how critical data and issues like data quality are to the overall governance of AI in financial services firms.

Data standards and regulation

70. As firms develop and evolve their data strategies to use of AI systems, there is an increasing call for the development and use of data standards specific to an AI context. While the financial sector already has data standards and regulation in place in the form of BCBS 239[40] and MiFID II,[41] additional incremental standards may be needed for most elements of traditional data governance, including data documentation, data quality, retention, privacy, sovereignty, and security. For example, the Alternative Data Council has started to produce standards for the purchase and use of third-party alternative data by investment firms.[42]

Box 4: BCBS 239

BCBS 239 (the Basel Committee on Banking Supervision's Standard number 239 on 'Principles for effective risk data aggregation and risk reporting') is a set of principles aimed at strengthening bank risk data aggregation and internal risk reporting. The standard, which was published in 2013 and came into effect from January 2016, applies to Global Systemically Important Banks (G-SIBS).

The standard details eleven key principles structured under three sections: [I] Overarching governance and infrastructure; [II] Risk data aggregation capabilities; [III] Risk reporting practices. The standard also has two additional sections describing supervisory principles and detailing timelines and transitional arrangements. More details on the principles can be found in Annex II.

71. Adapting BCBS 239 to include requirements on representativeness and data cards could be a way of advancing the data standards debate. Current BCBS 239 standards include requirements on accuracy, timeliness, flexibility, and completeness. However, they do not include standards on representativeness, which is very important when training AI models. They also lack metadata standards and this could be addressed by the development of data cards. A data card would be a short document that provides key information about the data, including the metadata, and presents it in a structured way. Data cards would be similar to existing AI model cards.[43] Adapting BCBS 239 standards in this way could be done for a single use-case first (e.g. consumer credit) as a way of piloting the idea.

72. Related standards may also be needed for the development and roll-out of technologies needed to integrate data processes with the AI modelling process. These are mainly techniques for sharing datasets while ensuring that private, personal data are protected. These techniques include approaches such as differential privacy, federated learning, and homomorphic encryption (see Box 5 below). Most of these are not unique to AI but they become significantly more important in the context of AI.

(40) Principles for effective risk data aggregation and risk reporting (BCBS).
(41) MiFID II is the 'Markets in Financial Instruments Regulation II', a European Union directive (2014/65/EU) for regulating financial markets. See Article 64 for specific data standards to help increase transparency in over-the-counter markets.
(42) Alternative Data Council – (FISD SIIA).
(43) Model cards for model reporting (Mitchell, et al. 2019).

Box 5: Privacy-preserving methods for sharing data

Differential privacy is a method for sharing characteristics of a dataset without disclosing information on individuals in the dataset.

Federated learning is a method of training models across distributed and/or decentralised networks using local data and without exchanging those data.

Homomorphic encryption involves the use of cryptographic techniques allowing users to carry out calculations on encrypted data without needing to decrypt them.

73. While achieving equal access to and fair pricing of third-party data is a reasonable aim, there are associated risks. For example, a push for equal access may disrupt the potential for markets to identify high-quality datasets and providers. Equal access to third-party data may also cause incremental risk in data handling as data are transferred between multiple parties in the value chain.

74. Lack of trust in third-party data quality is preventing some firms from using such data. Regulations that allow firms to trust third-party data would allow these firms to do more. Data federation, a database management process allowing multiple independent databases to function as one (such as the approach proposed by GAIA-X)[44], is a possible solution. Such an approach allows data sharing while having very strong governance allowing for access, traceability, and auditability.

75. Additional guidance on existing data regulation (both UK and international) could be useful. At the international level, there is value in international alignment for both data and AI governance. Coordination across industry and across financial services regulators, both domestically and internationally, could also be helpful. Instead of a more prescriptive approach to regulation of third-party data, a good alternative is the example of food labelling:[45] the vendor must supply information on the age of data, how representative the data are, when permission was last sought etc.

76. The creation and use of cross-firm datasets that support the public good and increase financial inclusion is a desirable goal. However, it can be difficult to achieve for a range of reasons, including: privacy and consent, commercial considerations, disclosure requirements, interoperability, data quality, and security. Cooperating on 'data for good' is nevertheless extremely valuable, as underlined by work in response to the Covid pandemic.

77. The development of standards for data cards (and model cards)[46] could be a useful initial step. In particular, developing a use-case specific data-card, together with a representativeness requirement coupled with existing BCBS 239 standards could form an updated data standard for that use-case. However, data cards are only a partial solution to some of the issues outlined above as their benefits tend to come at the end of an AI project. A more anticipatory approach to governance is key to managing AI risks.

Suggestions for good practice – Data

Firms should:

- Consider adoption and use of AI holistically, harmonising, where possible, data and model processes. In particular, they should aim to coordinate data management and strategy with AI management and strategy.
- Have processes in place for tracking and measuring data flows within, as well as into and out of, the organisation.
- Carry out regular data audits and assessments of data usage, including alignment with business objectives.
- Have a clear assessment of the value of the data it holds and uses, both for specific projects, and to the firm as a whole. This should inform the cost/benefit analysis of the AI project.
- Have clear understanding and documentation of the provenance of data used by AI models, especially in the case of third-party data.
- Have a clear understanding of the limitations and challenges of using alternative and/or synthetic data.

(44) What is Gaia-X? (Gaia-X European Association for Data and Cloud AISBL).
(45) Food labelling and packaging: Food labelling – what you must show (GOV.UK).
(46) Model cards for model reporting (Mitchell, et al. 2019).

Model risk

78. The third AIPPF meeting[47] and subsequent workshops focused on the model risk associate with AI; the challenges firms face when managing those risks; and emerging best practice to address them, both through current model risk management (MRM) practice and new approaches.

79. As the introduction chapter notes, mathematical and statistical modelling are not new to financial services, and neither are the risks involved. However, as this chapter explores, AI represents a step change for several reasons, which may amplify existing model risks and introduce new ones. That is why MRM is becoming increasingly important as a primary framework for managing AI-related risks in financial services.

Box 6: What do we mean by model and model risk?

There are many definitions of *model* and *model risk*. In this document, model refers to a theoretical construct that specifies the relationship between sets of input and output variables. One definition widely used in financial services is that in the Federal Reserve's Guidance on Model Risk Management (SR11-7).[48] It states that a model is a '*quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates*'.

Model risk, on the other hand, is the risk arising from use of models and can result, among other causes, from inadequate model specification, poor model implementation or incorrect model use. Although limited to capital models for banks, building societies, and designated investment firms, the EU Capital Requirements Directive (CRD IV, Article 3.1.11)[49] defines model risk as 'the potential loss an institution may incur, as a consequence of decisions that could be principally based on the output of internal models, due to errors in the development, implementation or use of such models'.

Key findings – Model risk

Most of the risks from using AI models in financial services are not new and can be caused by the use of non-AI models. What is new, is the scale at which AI is beginning to be used, the speed at which AI systems operate and the opacity or complexity of the underlying models. These, coupled with data risks, including the use of alternative, unstructured data, can create new challenges or amplify existing ones. Inadvertent risks can also emerge because there are unknowns with AI, especially when multiple models interact within a network.

Complexity is the key challenge for managing the risks arising from AI models. This includes complexity of the inputs (multiple input layers and dimensions); relationships between variables; the intricacies of the models themselves (e.g. deep learning models); and the outputs, which may be actions, algorithms, unstructured (e.g. images or text), and/or quantitative.

Explainability is a key issue with AI systems. Approaches to managing the issue should not focus solely on model features and parameters, but also on consumer engagement, and clear communications.

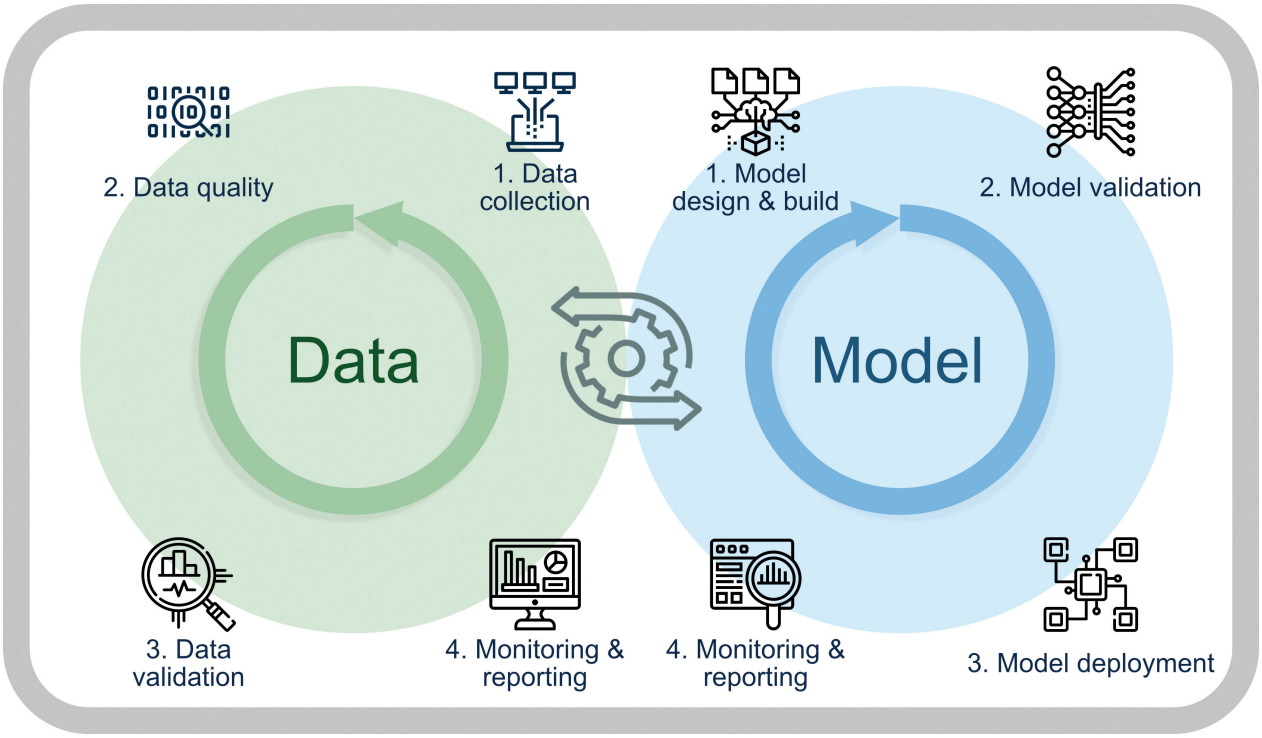
(47) Artificial Intelligence Public-Private Forum – Minutes 15 June 2021 (Bank of England).
(48) The Fed – Supervisory Letter SR 11-7 on guidance on Model Risk Management – April 4, 2011 (Federal Reserve).
(49) Capital Requirements Directive: Article 3 (European Banking Authority).

Identifying and managing change in AI models, as well as monitoring and reporting model performance, are a key part of ensuring that models behave as expected. This includes monitoring for changes in the functional form of a model as well as outputs of the model.

Model risk, governance, and lifecycle

80. Model risk exists at all stages of the data and model lifecycles (see Figure 4), and a firm's MRM processes should be aligned with those stages. While model risk is not specific to AI, the complexity, speed, and scale of AI models present new challenges and risks as well as amplifying existing financial risks. AI models can also introduce non-financial risks that are currently not as well understood, such as data privacy and protection, cybersecurity, and transfer learning (the use of knowledge gained in one domain to solving problems in a different but related domain).

Figure 4: Model and data lifecycles



Source: AIPPF.

81. One example of the new challenges arises from the dynamic nature of some AI models - their ability to learn continuously from live data and generate outputs that change accordingly. While dynamic AI models could outperform static models by adapting to changing data inputs (data drift) or changes in the statistical properties of the data (concept drift), firms should align governance processes and assessment of model risk to the adaptation cycle. Most AI applications currently used in financial services are static.

Box 7 – Drift

Model, data, or concept drift are common shifts in underlying data, variable relationships or statistical characteristics that lead to model underperformance and potentially inaccurate outputs.

Data drift refers to unanticipated changes in data inputs or structures that were not part of the model training process. *Model drift* on the other hand is the reduction in model performance due to changes in the relationship between input and output variables, potentially leading to inaccurate outputs and poor decision-making. And finally, *concept drift* refers to changes in the statistical properties of the target or output variables of the model.

82. MRM functions and processes within organisations must adapt to the various challenges that AI models introduce or increase. MRM for AI models requires, among other things: greater understanding of the use of hyperparameters; understanding and managing issues involving explainability and reproducibility; and data privacy and bias risks, which can also flow through to models and algorithms. These issues become even more challenging when using third-party models.

83. Model risk is broader than the model itself or the underlying algorithms. The model itself should not be seen as the sole source of risk; and certain types of data risk should be considered part of the MRM process. Risks can also arise in how a model is used and its business application, rather than the model or algorithm itself. For example, the size of a firm; the number of customers affected by the model; and customers' understanding or perception of the technology could also have an impact on model risk.

84. Reporting model risk across model types and use-cases and integrating with other risks across a firm can be very difficult, especially if this is not linked to underlying business risks and outcomes. It is useful to think of model risk as a residual risk or root cause of other risks (credit, liquidity, reputational). Both data risk and model risk can be aggregated when put in the context of an operational risk framework and linked to real business outcomes.

85. The Bank and the FCA should consider creating model risk regulations such as the Federal Reserve's SR11-7.[50] Some UK banks have used the SR11-7 principles as well as the PRA's Model Risk Management Principles for Stress Testing[51] as templates, to help with the lack of clarity in the UK guidance.

Complexity

86. AI models in financial services are becoming more complex, as are wider data modelling and data analytics methods. Smaller firms, especially recent entrants, can sometimes use more complex AI compared to larger firms as they are not constrained by legacy systems. In some cases, they may also be incentivised to use more complex AI models to compensate for a potential lack of data.

87. Complexity of the underlying AI models is the key MRM challenge. The increasing complexity of both model inputs and the ways in which a model works means that traditional MRM may become less effective. Monitoring outputs and performance may make more sense for AI MRM, rather than the more traditional MRM focus on assessment of inputs. But there are challenges with this approach since the outputs themselves are also becoming more complex.

88. Reproducibility (i.e. the principle that model development should be documented in such a way that the process can be repeated with identical results) is an important consideration, especially if customers ask a firm about a decision at a later date. But the scale of AI and data being used by firms can also pose a challenge, since it is not clear what data, models, and other metrics should be logged (for example: test data, training data, live business data, source code, explainability metrics, etc.), and for how long (weeks, months, years perhaps), all of which come at a cost to the firms. Models may be stochastic and hard to predict, which affects reproducibility.

89. There are benefits to a tiered, risk-based approach to MRM that would account for an implicit trade-off between the risks an AI model poses to an organisation and the complexity it is willing to bear. Although some models can be highly complex, they can also be used in low materiality applications. Similarly, high materiality applications in financial services tend to use less complex models because explainability for the relevant stakeholders (consumers, internal compliance, regulators, etc.) is very important.

90. Model complexity is not a new issue and existing IT complexity (and dependency) can, for example, be partly managed through microservices architectures,[52] with techniques such as complex unit testing.[53] This could provide a starting point for managing some of the issues around AI complexity. There is also a need for systems-level testing of more complex models. The more traditional, linear, sequential approach to model development and production may not be appropriate for AI models. The maturing of MLOps (the use of ML to automate operational processes), is

(50) The Fed – Supervisory Letter SR 11-7 on guidance on Model Risk Management – April 4, 2011 (Federal Reserve).
(51) PRA SS3/18, Model risk management principles for stress testing (Bank of England).
(52) Microservices architecture (Wikipedia).
(53) Unit testing (Wikipedia).

leading to the application of continuous ML (analogous to continuous delivery or continuous deployment) for AI processes with the focus on orchestration, testing, and monitoring.

91. Lastly, it is likely that firms will use more complex AI models for material use-cases in the near future as they become more comfortable and proficient with the use of AI. The challenge for MRM is what to do if and when those models do not perform as expected, or the outputs of those models deteriorate beyond an acceptable risk tolerance, or indeed, to understand how these complex models shape outcomes.

Explainability

92. One of the defining aspects of some AI models, such as deep neural networks, is the lack of clear understanding of their inner workings i.e. the black-box problem. To address this, there are increasing calls for more clarity on what level of explainability or interpretability is necessary, particularly for regulated activities.

93. What is meant by an 'explanation' depends on the context. The requirement for a good external explanation for a customer is different to that for internal explanations for data scientists or non-technical executives. These kinds of explanations can be so different that they may need separate processes for generating them. Even if there were no requirement to provide explanations to customers, firms would need to produce explanations for internal use to ensure better understanding of the AI system's workings and capabilities.

94. There is a difference between regulating models themselves with a high-level of explainability versus treating them as black-boxes and regulating their inputs, outputs, and outcomes. Which approach is more appropriate depends on the context and materiality of the use-case and, potentially, the regulatory context. Having clear guidelines on appropriate degrees of explainability for specific use-cases could increase confidence when using the technology in financial services, but it could equally hamper desirable innovation. Where explainability is not possible, the ability to explain the safeguards that are put in place to protect against negative outcomes should be considered. Striking the right balance is a key consideration for regulators and policy makers, as well as firms and third-parties.

95. It could be argued that the focus should be on the customer experience. When looked at in this way, explainability, while still important, becomes part of a much broader requirement on firms to communicate decisions in meaningful and actionable ways. From this perspective, the focus is not just on model features and important parameters, but also on consumer engagement, and clear communications.

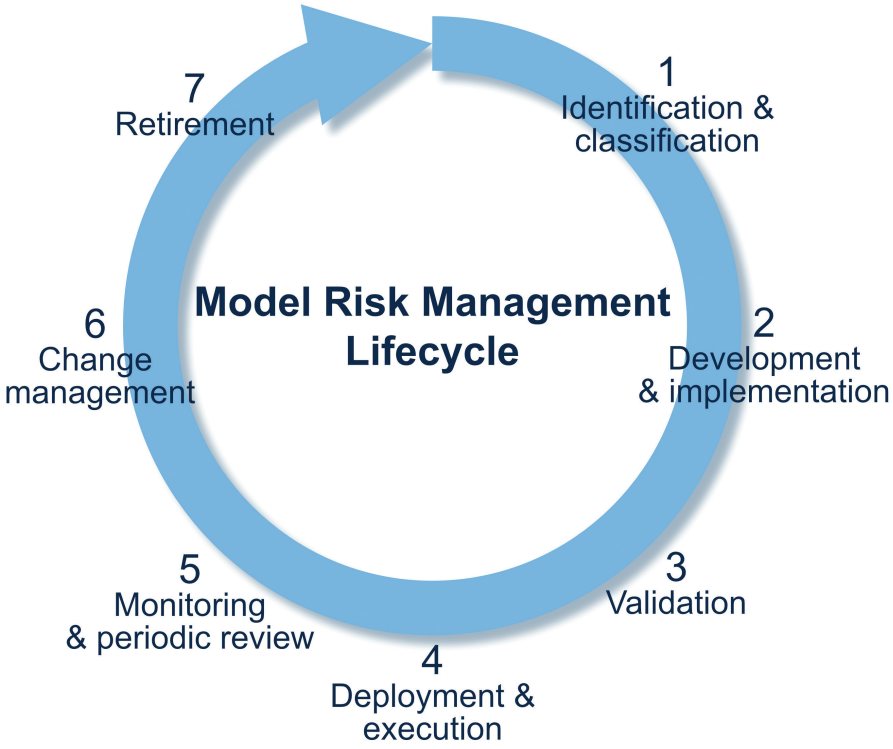
96. While there may be many design principles and guidelines within firms' governance procedures, these are not usually communicated to end-users. Communicating internal principles more effectively could be a useful adjunct to model explainability.

Change management, validation, monitoring, and reporting

97. Identifying and managing change in an AI model is one of the key MRM lifecycle processes (see Figure 5). A material change in an AI model may require going through the entire MRM lifecycle and governance process, which is time- and resource-intensive. So developers and firms sometimes put in place high thresholds for significant model change. Firms are able to do this because there is no clear guidance (or indeed conceptual clarity) on what constitutes a material change. This situation is made worse because, while AI models can have lots of small and incremental changes, it is not clear when these become a material change (this is analogous to the Ship of Theseus parable).(54)

(54) Ship of Theseus (Wikipedia).

Figure 5: MRM lifecycle



Source: AIPPF.

98. A vital part of MRM is identifying and managing model drift and defining when a change of the model's data or parameters might constitute a new model. For example, if the functional form of a model (e.g. the number of parameters) changes or the outcomes change significantly then that would constitute a new model which needs to be re-evaluated. This is especially the case with reinforcement learning models that can change their behaviour over time.

99. Current change management processes within firms may not be suitable for rapidly changing AI models since the time it takes to complete the change process undermines going through the process in the first place, even if the firm determines that a material change had occurred. Firms' change management processes have to adapt to meet the need for faster re-training and validation. For example, validation tends to happen periodically (often quarterly or annually) and most firms currently use a waterfall (sequential) approach. This process does not work for fast-paced, continual updates with AI models.

100. There is an increase in the importance of monitoring for AI systems but this is not always considered enough of a priority, especially during model validation stages. The validation step itself should ensure that there is appropriate monitoring in place for the model, which includes oversight of performance by First-Line-of-Defence (1LOD) and Second-Line-of-Defence (2LOD) teams.

101. Two of the key documents from an ongoing validation and monitoring process are: (i) a monitoring plan covering aspects like the frequency, metrics, and mitigating actions; and (ii) a change plan covering the different elements of the model infrastructure that are allowed to adapt over time, setting guardrails and the extent to which they can adapt, and a documentation process for any change.

102. There is a question on how firms could best monitor and document the various harms and risks raised by AI. While automated monitoring and change management processes increase efficiency, firms should understand clearly who is responsible for the automated pipeline and, specifically, the monitoring aspects. Clear lines of accountability are crucial for the automation of AI model monitoring and change management processes.

103. Finally, differentiating between validation and ongoing monitoring is important for managing AI model risks. Especially because the latter may need to be real-time for certain AI models and conducted by 1LOD functions. Traditionally this was done periodically by the Third-Line-of-Defence (3LOD) and audit functions – an approach that may no longer be effective.

104. Relevant thresholds and metrics for effective AI risk monitoring and change management are context and use-case specific. The metrics should be relevant to the business impact of the model. Similarly, understanding business outcomes and metrics may help inform the choice of model. For example, firms could ask themselves whether less complex or non-AI model achieve the same outcome, reducing the need for complex models in certain circumstances.

105. One approach to the challenge of monitoring and auditing third-party models is to shift responsibility from vendor to client via contractual agreements. This occurs in most cases and happens because vendors are unlikely to accept liability for clients' use of their models. Moreover, many clients train third-party models on their own proprietary data, which can affect liability. However, there may be ethical concerns with this approach, in particular from civil society which wants technology providers to assume greater responsibility for the models they put into commercial use. Similarly, the draft EU regulation on AI proposes to hold both the vendor and client accountable for credit-scoring AI models.

106. The use of third-party models and data has increased significantly over the last few years, accelerated by the pandemic(55) and the need for cheaper, easy-to-deploy AI solutions. However, while there are clear benefits to using third-party models and data, there are many associated challenges, particularly in ensuring full due diligence of the model that is being outsourced.

107. A further question is how firms and their senior managers could comply (and demonstrate compliance) with their responsibilities when relying on third-party service providers. Contractual relationships can set in place obligations for responsible practice on both sides. These should include regular monitoring for performance and outcomes, both internally and with the third-party provider (i.e. a 'shared responsibility model').(56)

Backup and remediation

108. Given the potential model risks associated with AI, backup options and remediation actions need to be a consideration before the model is put into production. Such an approach will enable firms to respond appropriately and in a timely manner if an AI model performance or output deteriorates beyond the accepted risk threshold, which can help manage risks.

109. As with the monitoring and change management processes, backup options and remediation actions should be context specific and linked to business outcomes. For example, an Optical Character Recognition AI model may be used to analyse customer documents to identify specific numbers or fields of information. If that model fails, will the firm still be able to process customers that day, does it have enough human agents to do the job while the AI model is not in operation, is there another model that can be used instead?

110. In terms of potential solutions, one approach is to quickly retrain and replicate the model. However, retraining is hard, time consuming, and can involve a long governance process. So it is useful to have a simple backup or shadow that firms can switch to and use while they retrain the model, with the promise of the updated AI model to follow once it has gone through the appropriate checks.

111. As with many other issues in this report, operational and physical resilience is not AI-specific (it applies to any technology system within firms) and it is already an issue for firms, rather than something to consider in the future. However, the one area where AI differs is the resilience of the training process, which is not currently treated with the same operational resilience requirements. This means risk teams are not checking that the retraining is on distributed nodes or being chaos-tested, which has caused some problems for some firms.

(55) The impact of Covid on ML and Data-Science in UK banking (Bank of England).
(56) SS2/21 'Outsourcing and third party risk management' (Bank of England).

112. There could be merit in applying operational resilience requirements[57] and processes to the AI retraining process. Combining the data science, software engineering, and MRM teams would likely help bring about that change. It would also improve the overall incident management process, which tends to be less streamlined because the software engineering and data science workflows are separate. Ideally, firms could develop a hybrid skillset within teams but it is difficult to find and hire the relevant individuals.

113. It can be easy to get distracted by the technology aspects but the remediation issue is largely about process and knowing the right people to address the problem. This can make the difference in terms of the time it takes to address any issues and the impact on consumers. Ultimately, before a model is put into production, firms need to discuss with the relevant stakeholders what could go wrong and, if something does go wrong, who is responsible and what actions they need to take.

Box 8: Networks and other systemic risks

While some of the systemic risks associated with increased use of AI in financial services have been well-documented, others are not so well-understood or researched. Within the former are the potential increase in herding and procyclical behaviours in financial markets. These may, at the extreme, lead to flash bubbles or crashes when considering trading and sectoral, demographic, or regional concentrations when considering credit. There could be an increase in cybersecurity risks to and from AI models which itself could pose systemic risks.

The potential for networks or clusters of AI models to have a significant and unpredictable impact on markets can become a systemic risk with AI's wider adoption and use. Inadvertent risks can emerge because there are many unknowns with AI, especially when multiple models interact within a network. For example, the interconnected nature of models, where the outputs of one model become the input of another model, means there is also a risk that the reward scenario for one model could drive unwanted behaviour in another model. This may, in turn, have negative implications for consumers, firms, and the financial system as a whole.

The use of decentralised and networked AI systems increases risks from network structure. For example, some nodes may become important purely because of their position in the network rather than any innate properties. Networks may also exhibit emergent behaviour (when a number of simple entities operate in an environment and form more complex behaviours) both at the micro and macro levels. Emergent behaviour is very difficult to predict and its implications difficult to project and analyse.

In addressing unintended and systemic risks, the key question is how much emphasis there should be on a robust system-wide monitoring framework both before deployment and on an ongoing basis. Such a wide array of effects arise from putting AI models into production that a small set of metrics is unlikely to give a full picture. Instead, firms need to think about the wider impact of models outside the initial field of application or business area, including on markets and consumers.

Suggestions for good practice – Model risk

- Firms should have:
  - A documented and agreed AI review and sign-off process for all new applications.
  - A complete inventory of all AI applications in use and in development.
  - Clearly documented methods and processes for identifying and managing bias in inputs and outputs.
  - Regular assessments of AI application performance, including potential external impact.
  - A clear explanation of AI application risks and mitigation (to be updated as appropriate).
  - Documentation and assessment of AI application inputs, including data quality and suitability.

(57) PS6/21, CP29/19, DP1/18, Operational Resilience: Impact tolerances (Bank of England).

- An appraisal process for explainability approaches (for internal, regulatory, and consumer use – potentially three different approaches).

- The benefits AI brings should be commensurate with the complexity of the system. Firms should be able to demonstrate full understanding of why they are using an AI application compared to something that is simpler and easier to understand that produces similar outputs.

- In addition, current MRM practices may not fully cover consumer harm and risks. MRM should try to measure the impact on consumers (for example, those that are denied access to credit) and help manage those risks as well as the potential impact on markets. MRM could also involve a communication step, where consumers are informed when and where an AI model is involved, and given all the relevant information.

Governance

114. As the fourth AIPPF meeting[58] emphasised, governance is crucial to the safe adoption of AI in financial services. It ensures accountability and puts in place the set of rules, controls, and policies for a firm's use of AI. Good governance can ensure effective risk management and help address many of the data and model-related issues raised in the previous chapters. On the other hand, poor governance can increase challenges and produce risks for consumers, firms, and the financial system.

115. As with other aspects of using AI in financial services, ensuring appropriate and effective governance can be difficult to achieve. One reason is the incremental capacity for autonomous decision-making, which means AI can limit or even potentially eliminate human judgement and oversight from key decisions. Another reason is that AI systems touch upon various existing governance functions, which can make it difficult to have clearly defined lines of accountability and creates challenges for the 'three lines of defence' model. The skills gap that exists at all levels tends to exacerbate these challenges.

116. There are also wider issues when it comes to addressing issues like bias and fairness, and understanding the role of auditing and certification regimes. This chapter explores these and other challenges, as well as some potential ways to address them.

Key findings – Governance

A key characteristic of AI systems is their capacity for autonomous decision-making. This can have profound implications for how to govern the technology and its outcomes, including ensuring effective accountability and responsibility.

Existing governance frameworks and structures provide a good starting point for AI models and systems, partly because AI models will invariably interact with other risk and governance functions. The most relevant in financial services are data governance and MRM frameworks, as well as operational risk management.

Governance of AI in firms is more effective if it includes diversity of skills and perspective, and if it covers the full range of functions and business units. This type of cross-functional approach can help manage the complexity of AI systems and their associated data challenges.

Governance of AI systems should be aligned with the risk and materiality of the use-case, even when existing governance frameworks are applied to AI. Where possible, firms should leverage and adapt existing governance structures to manage AI including data and MRM frameworks.

Transparency and communication are key elements of AI governance.

Standards for AI governance should be set by a centralised body within firms. Overall responsibility for AI could be held by one or more senior managers, with business areas being accountable for the outputs and adherence to the governance standards.

Firms should ensure that there is an appropriate level of understanding and awareness of AI's benefits and risks throughout the organisation.

A key focus of AI regulation should be on how AI affects decision-making. Regulators should provide greater clarity on the types of outcomes they expect for AI governance and controls. However, enforcing metrics around outcomes may prove to be challenging in practice.

(58) Minutes of the Artificial Intelligence Public-Private Forum - 1 October 2021 (Bank of England).

Establishing an auditing regime for AI practitioners and professionalisation of data science would help foster wider acceptance of and trust in AI systems.

Governance frameworks and structures

117. Existing governance frameworks and structures provide a key starting point for AI models and systems. This is in part because AI models almost invariably interact with other risk and governance processes. The most relevant in financial services are data governance, MRM frameworks, and operational risk management. Many of the risks from AI models also apply to more traditional models with existing governance structures to manage those risks and issues. Therefore, where possible, firms should use and adapt existing governance frameworks to manage the novel challenges of AI. However, there are some aspects to the use of AI that may require new risk management and governance frameworks.

118. One approach could be to combine existing governance frameworks into one overarching AI governance framework. MRM, which is the primary governing framework for AI in most banks, could be combined with data management to form a single framework to address AI-related issues more holistically. It could also be beneficial to develop a set of AI risk principles and map them to existing risk frameworks. This would enable firms to identify and focus on potential risks not already covered, including areas where staff need to be trained. This is particularly relevant as AI governance and issues like ethics will require a broader set of skills, experiences, and backgrounds rather than a narrow focus on technology, risk management, and compliance.

119. Another approach is to develop a cross-functional body, or 'council', that includes compliance, audit, research, data, MRM, and other areas. This would help address the need for a broader and more diverse set of skills, which could ensure risks are not overlooked. The council could be chaired by a relevant executive recognised by the Senior Managers and Certification Regime[59] (SM&CR) but involve checks and balances from all relevant risk and compliance areas. One potential challenge with this approach is that it could create more onerous processes and communication issues if there is no shared lexicon or common understanding of key concepts. However, there may be ways to streamline some of the processes.

120. Whatever the approach, governance frameworks and processes should be tiered and aligned with the risk and materiality of the use-case. For example, high-risk and high-impact AI use-cases (like consumer credit) will likely require more due-diligence and, therefore, time and resources. Whereas low-risk use-cases (like chatbots) may involve less due-diligence and a more streamlined approach. It is important to incorporate customer outcomes and the impact on customers into such risk-based AI governance frameworks. The rapid pace of development in AI also means that any governance frameworks (including the associated standards and risk ratings) must be flexible, reviewed regularly, and refreshed when necessary. This implies two levels of governance: (i) a strategic-level that develops standards for the entire firm centrally; and (ii) an execution-level that applies the standards on a use-case basis, distributed across the firm.

121. Governance frameworks should also seek to deliver a safe environment for testing AI models. One of the factors that can inhibit innovation (across all industries) is an imbalance in incentive structures, which heavily penalise mistakes and result in a high cost of failure. This is particularly true for sensitive datasets, which can provide some of the biggest benefits if used correctly and ethically but carry the highest costs when something goes wrong. The risk-based approach described earlier could help enable safe testing and further innovation.

122. Another way to encourage innovation is to automate certain governance processes, such as documentation generation or information/evidence collection. For example, there are techniques for embedding automation to flag potential risks and feed information into a central system to produce documentation. Similarly, many aspects of model testing can be automated and set to occur on a regular basis. In certain use-cases, such as fraud detection and AML, firms can automate the generation of synthetic data to check for issues like bias and stress-test the models. Overall, automation allows relevant information or documentation to be embedded in the development lifecycle, rather than being separate activities undertaken at a later stage, so speeding up processes such as model approval and monitoring.

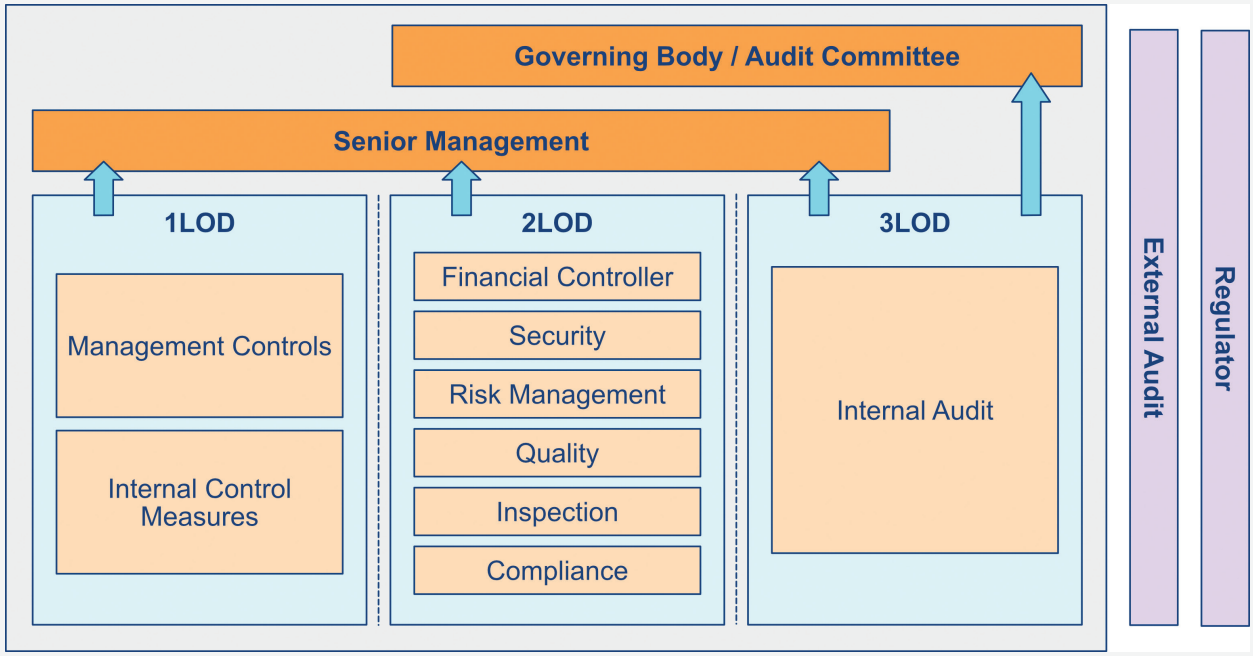
(59) Senior Managers and Certification Regime (FCA).

123. As noted in the Model Risk chapter, financial services firms often use the three lines of defence model as part of their overall risk management and governance framework (see Figure 6 and Box 9 below). However, the use of AI systems can present a number of challenges. For example, AI systems demand a more integrated approach encompassing 1LOD data scientists, developers, and model users, as well as 2LOD risk, compliance, and oversight functions. This is because AI models can be more dynamic and some can adapt and learn continuously, which makes modelling, development, monitoring, and validation an iterative cycle. At the same time, there is an even greater need for challenge from and independence of the 2LOD to ensure appropriate risk management.

Box 9: Three lines of defence

One of the key governance and risk management concepts in financial services (and many other sectors) is the three lines of defence model.(60) In this framework, the business areas are the 1LOD, independent risk management units are the 2LOD, and internal audit is the 3LOD.

Figure 6: Three lines of defence



Source: AIPPF.

124. Striking the right balance between the scope of different lines of defence can be challenging. This is often made more difficult by the lack of relevant model risk and validation skills in the 1LOD teams, and lack of data science and AI skills in the 2LOD functions. For example, the move towards automation also changes the required skill-set in the 1LOD. That is why 1LOD teams may need software engineers as well as data scientists, helping to integrate automated monitoring controls with the existing MRM framework.

125. One way to address the skills gap and split in responsibilities is to provide more training and promote closer collaboration between the 1LOD and 2LOD. In particular, between data science teams in the 1LOD and 2LOD MRM teams. However, this requires significant investment by firms, may lead to a realignment of staff incentive structures and erode the independence of the 2LOD.

126. An alternative to closer collaboration between the 1LOD and 2LOD, is to integrate various 2LOD functions and rely less on one specific function, such as MRM. Most 2LOD functions are small teams that may not be able to manage the scale of AI models. For example, AI presents a range of risks (not just model risk) and it is unrealistic to think that MRM 2LOD functions are experts across all risk areas. So AI risk management could be shared between MRM, data, cyber, IT, compliance, and other 2LOD risk functions within firms.

(60) Regulatory expectations (Bank of England); Internal audit: three lines of defence model explained (ICAS).

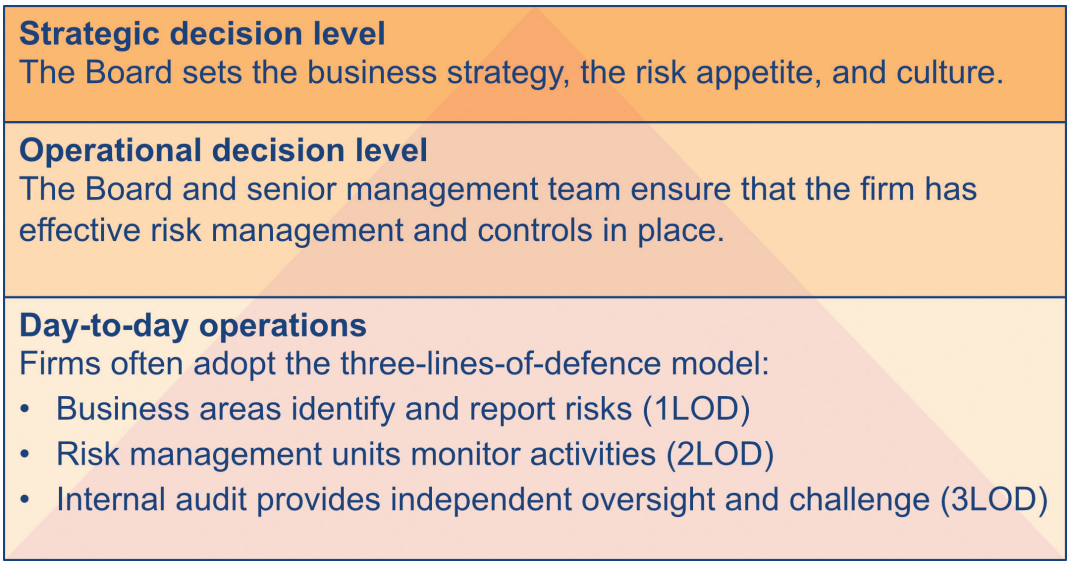
127. An enterprise-wide approach to risk management could help address the complexity and scale of AI. For example, a cross-functional or enterprise-wide 2LOD function could capture the outputs of all models in all business areas, including models developed internally and those provided by third-parties. This would enable the 2LOD function to have a more complete and aggregate view of potential risks across the organisation, including model, operational, credit, market, traded, and cyber risks.

Accountability and responsibility

128. Accountability, responsibility, and informed decision-making are central themes to any discussion on AI governance. This is a common feature of the numerous AI governance principles that have been produced around the world for financial services and other sectors,(61) many of which have specific principles for accountability.(62)

129. Whether firms adapt existing governance structures or establish new frameworks, they need to clearly define the relevant roles and lines of accountability at all stages of the AI governance hierarchy (see Figure 7). This includes allocating responsibilities across the firm, from the design and development of AI models, to the business areas that use models, the compliance teams that oversee the risk management of those models, up to senior managers and Board members. Exact lines of responsibility and accountability within firms may differ depending on the maturity of AI use-cases as well as the firm's size and type. This makes it unlikely that a single, common approach will be suitable for all financial services firms.

Figure 7: AI governance hierarchy



Source: AIPPF.

130. One of the key challenges for AI governance is whether an organisation should centralise or decentralise responsibility for AI. A key question for all firms is who should ultimately be responsible for AI, including under the SM&CR, and whether this should be a single individual (e.g. Chief AI Officer) or shared between several senior managers (e.g. Chief Technology Officer, Chief Data Officer and Head of MRM). Whatever approach firms take, there should always be clear lines of accountability and responsibility for the use of AI at the senior managers and Board levels.

131. It may be helpful to have a centralised body responsible for the firm's AI governance policy. The centralised body should have a complete view of all AI models and projects to enable it to set the standards or policy for managing AI models and associated risks. Centralised accountability for setting standards and policies can be more effective, meaningful and comprehensive, compared to multiple business areas setting different standards within a firm. A centralised body could also play a role in providing education, training, and relevant information on AI governance throughout the firm.

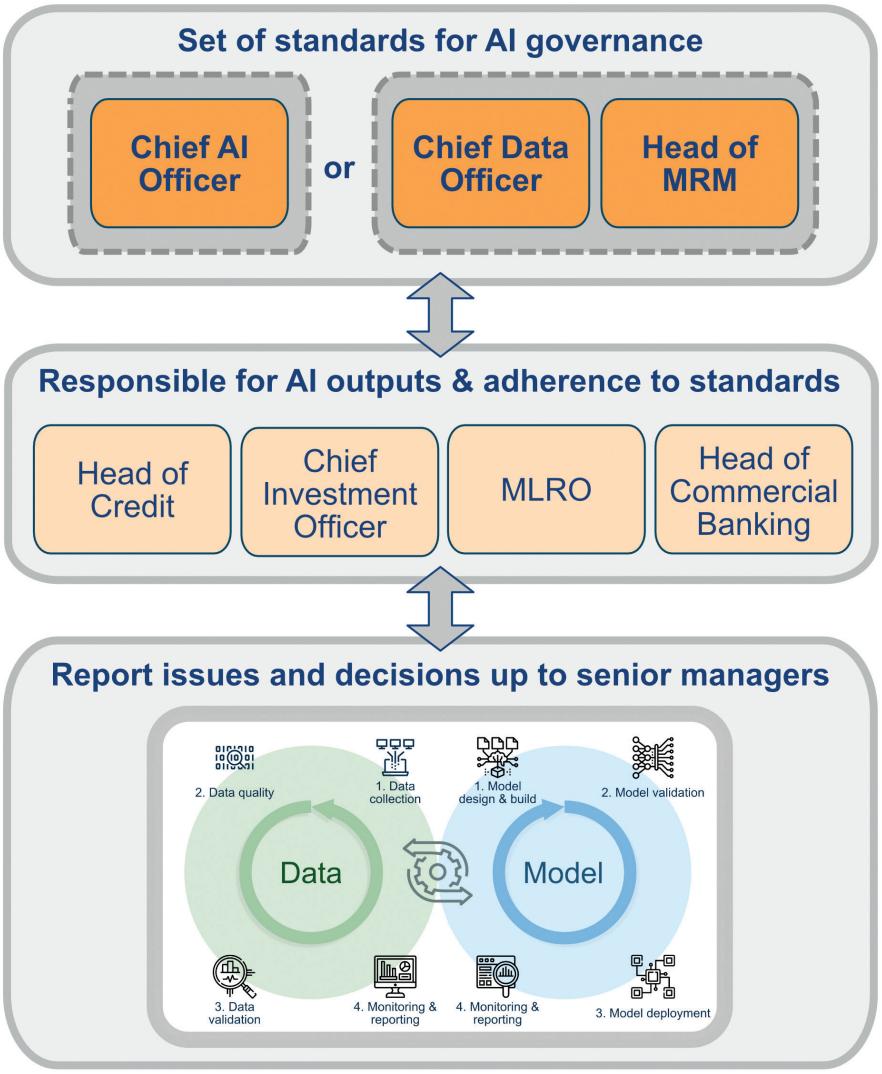
(61) AI Ethics Guidelines Global Inventory (Algorithm Watch).
(62) FEAT Principles Final (Monetary Authority of Singapore).

132. In contrast to a centralised body that sets the governance standards, responsibility for AI models themselves could sit at different levels of the firm in a decentralised or federated manner. The business areas that use AI models should be accountable and responsible for the outputs as well as adherence to and execution against the governance standards. At the senior managers level, that means the accountable executive for the business area should be responsible for the decisions made using AI models and adherence to the governance standards and process (e.g. Chief Investment Officer or Head of Credit).

133. *Reasonable steps* is a key concept in the SM&CR and the FCA Code of Conduct,(63) and could be extended to the use of AI. While assessment of what constitutes *reasonable steps* is a judgment-based process, it could include: having an ethics framework and training in place, maintaining documentation and ensuring auditability, embedding appropriate risk management and control frameworks, a culture of responsibility (ethics, governance, inclusion, diversity, and training), clear lines of oversight, reporting and accountability between AI teams etc.

134. Top-down and bottom-up governance approaches are not mutually exclusive and can also work well for AI. A production-line approach to bottom-up AI governance could alleviate many issues, including questions of where responsibility should sit for each stage of the AI lifecycle, particularly the split in accountability and responsibility between AI developers (e.g. model build teams) and implementers (e.g. business leads). This approach can provide a governance mechanism linking small decisions up to senior managers, including the decentralised layer responsible for AI models and adherence to governance standards, and the central body responsible for setting the overall AI governance standards and policy (see Figure 8).

Figure 8: AI accountability and responsibility



Source: AIPPF.

(63) COCON 4.2 Specific guidance on senior manager conduct rules (FCA).

Transparency and communication

135. Transparency and communication are also key elements of AI governance. There are two distinct audiences for transparency: (i) model developers, compliance teams, and regulators, and (ii) consumers who may be affected by model decisions. For the first group, firms should be able to provide accurate information about the decisions taken and assurance that recommendations and decisions are reliable. This may involve the use of explainability techniques like Shapley values(64) and LIME plots.(65) Although it is worth noting, no particular technical standard has proved superior to another and solutions are very context-dependent.

Box 10: SHAP & LIME

SHAP – Shapley Additive Explanations

SHAP is based on the Shapley value, which is a game theory method for assigning payouts to players depending on their contribution to the total payout. SHAP helps quantify the contribution that each feature brings to the model prediction by assigning each feature an importance value and so help provide an explanation of the importance of features or inner workings of the black box.

LIME – Local Interpretable Model-Agnostic Explanations

LIME approach uses local or surrogate models to estimate effects. These local or surrogate models are trained to approximate the predictions of the underlying black box model. Instead of training a global surrogate model, LIME focuses on training local surrogate models to explain individual predictions.

136. For the second group, consumers who may be affected by model decisions, transparency may be as important as communicating how specific decisions are made. Consumers should be told when a model is being used to automate decisions. The information asymmetry between customers and firms could also present challenges to ensuring appropriate levels of transparency. A full explanation of AI decisions in terms of explainability measures such as SHAP or LIME may not be needed or useful, but customers could be told what data were used and the most important features that led to the decision. Best practice on transparency for consumers when AI is being used for decision-making is still evolving but standardised approaches and a responsive attitude may lead to better consumer outcomes.

137. Resilience of internal systems is a further consideration when thinking about transparency. AI systems can be quite complex and some of the elements, though not necessarily the AI component itself, may be compromised and open to cyber-attack if too much detail of the inner workings are disclosed.

138. A particular case is the use of AI in financial crime and AML. Revealing too much detail of how those models work could expose them to potential attacks, rendering them ineffective. Resilience issues in credit lending, for example, may also lead to systemic risks. Firms need to be increasingly aware of attacks, sometimes based on public disclosures. Some AI models may also not be very stable. For example, some face recognition models can be fooled by changing a few carefully chosen pixels, and these can be attacked relatively easily.

Culture and skills base

139. Ensuring that those responsible for the use of AI in a firm have the relevant skills and knowledge is a key challenge. Accountable executives and senior managers should have an appropriate understanding of data, algorithms, models, and risks to fully consider any trade-offs. One of the main barriers here is the lack of necessary skills and buy-in from senior managers and business areas.

(64) A Unified Approach to Interpreting Model Predictions (Lundberg and Lee, 2017).
(65) "Why Should I Trust You?": Explaining the Predictions of Any Classifier (Riberio, et al. 2016).

140. Adequate AI oversight requires 1LOD model implementers in business areas to have sufficient understanding of the models used. On the other hand, 2LOD functions need clear parameters for performance and outcomes of AI models, including the ability to measure and monitor them, as well as adequate training to identify model drift or incorrect outcomes and to overrule decisions where appropriate. Diversity of skills and background should be a key consideration when considering firm culture: greater diversity will lead to a richer set of questions, more effective challenge and oversight.

141. More generally, the use of AI and associated skills gaps are shifting power relationships between individuals, groups, and institutions. In some cases, these shifts involve the creation of new power relationships and in others they can widen existing misalignments. For example, the relationship between 1LOD technology teams developing AI models, and 2LOD and 3LOD departments involved in the assurance of those models (such as audit and compliance teams) has shifted. This is because AI is more complex and opaque than traditional techniques, which means the 2LOD and 3LOD assurance teams' ability to monitor, validate, and effectively understand AI models is reduced relative to static models.

Bias, fairness, and ethics

142. Although problems around bias and fairness are not intrinsic to AI models, they are potentially worsened by the ability of these models to extract (often hidden) patterns and relationships or to exploit existing biases. Biases may, for example, be embedded in aggregated data or processed through proxy features. The use of AI models may compel institutions and users to define fairness in mathematical terms and code it into their models, in so far as that is possible. There has already been much work on treating customers fairly in financial services, and it may be possible to expand and build on it to incorporate AI.

143. While there is a plethora of principles and guidance on AI ethics from public and private sector sources, operationalising them and measuring implementation is challenging. It requires a clear understanding of the underlying principles and intended outcomes at every level of the organisation as well as very clear definitions at every stage of the process.

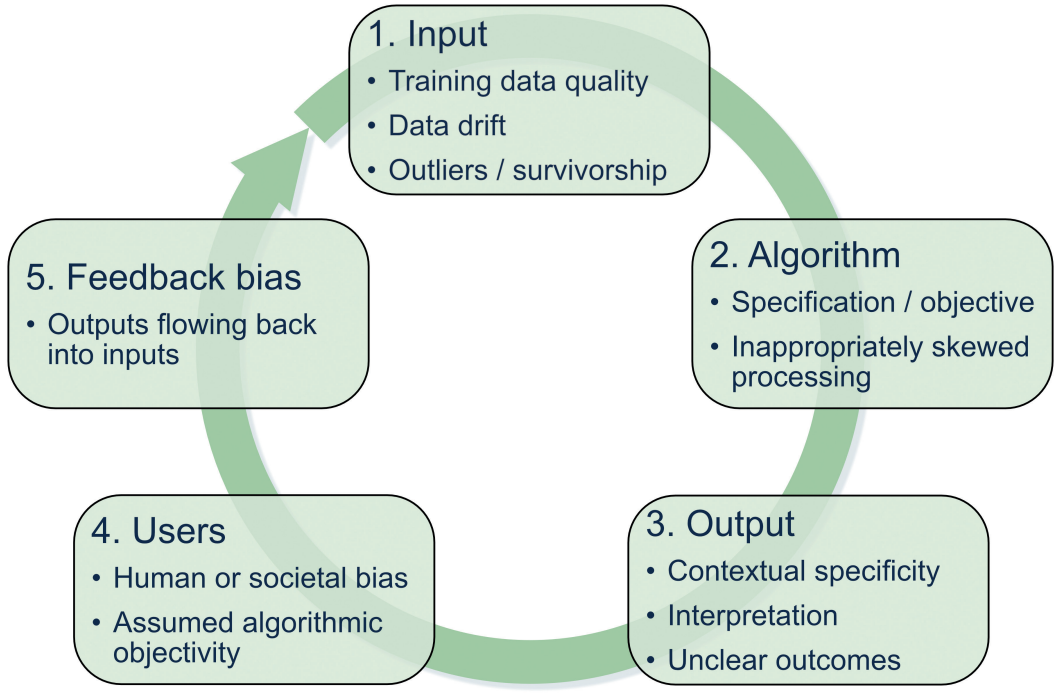
144. Conflation of the terms 'fairness' and 'bias' can be a challenge. There are many definitions of fairness; which one to use is often a business decision, reflecting the goals and outcomes of the specific business area. Businesses may also be uncomfortable discussing fair outcomes and may not have as clear a philosophy on fair outcomes compared to fair processes.

145. Change management in organisations is an important aspect of fairness and bias considerations. Specifically, creating the right environment for conversations on ethics takes time and requires buy-in from senior leadership. The right skillsets are needed to discuss ethics and fairness. A multidisciplinary and diverse approach is helpful, including a culture that allows internal challenge. Again, the use of third-parties can create additional challenges, since vendors and customers can have different ways of assessing fairness, including for data inputs or model design. Similarly, what may be considered fair can change as society and demographics change over time.

146. A further question is whether fairness and bias apply to businesses and corporate entities, rather than just to individuals. Do obligations of fair treatment by corporate institutions apply to interactions with corporate clients or only to retail clients? Within certain legislation and regulation, small and micro-businesses have similar rights to natural persons. How concepts of fairness apply beyond natural persons could, for example, have implications for lending to small businesses versus large businesses.

147. There also needs to be a distinction between different types of biases (see Figure 9). Although human biases are at the heart of many of these, it is important to understand how AI systems may exploit them. The first broad category is bias that is inherent in the data used to train AI models. Secondly, there is bias that emerges from the models. Many biases are intended and central to business considerations (e.g. higher insurance premiums for higher risk businesses). It is the unintended biases and unintended outcomes that should be of concern. So it is important to have frameworks and decision-making processes that distinguish between 'good' and 'bad' bias.

Figure 9: Different types of data bias



Source: AIPPF.

148. Defining a list of features that should be excluded from datasets and models is not realistic. The burden remains in proving that there are no proxy relationships between variables and known protected characteristics that result in bias. The question then arises on where the responsibility to remove bias should lie: with firms, which need to be mindful when building and deploying models; with regulators, who would need to understand and keep an updated list of features that should be excluded; or both.

149. Regulators could assure themselves that the firms they supervise are evaluating bias and fairness appropriately by: (i) defining a set of criteria for measuring bias in models; and (ii) setting criteria for evaluating bias and fairness of their models throughout the development lifecycle, from both the technical and business sides. There is a wider social dimension to the bias and fairness discussion, raising questions on the extent to which firms have a social responsibility to identify issues like individuals or families at-risk, including those identified by AI models.

150. There is a risk to being bound to what is known today and losing sight of what changes might come in the future, not only in how data are classified, but also on how data are collected. For example, in the past, people only used two genders to refer to themselves, but changing social norms mean that there are now alternative ways to express self-identity. As another example, there are geographical areas where technology is not as widespread, and relying entirely on data sourced from mobile devices, for example, would be misleading because data collection is not representative.

151. Another issue is who is 'in the room' when the AI models or use-cases are created; it is important to ensure that there is a diverse team and a diverse culture that permits internal challenge to critically review and challenge algorithms and data from an ethical standpoint.

152. The EC 'Ethics Guidelines for Trustworthy AI'(66) raises several related points. The Guidelines suggest that Trustworthy AI '*has three components, which should be met throughout the system's entire life cycle: (i) it should be lawful, complying with all applicable laws and regulations; (ii) it should be ethical, ensuring adherence to ethical principles and values; and (iii) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm*.'

(66) Ethics Guidelines for Trustworthy AI (EC).

Algorithm auditing and certification

153. Auditing and certification of AI models could help to address many of the governance challenges, as well as address the wider questions around transparency and explainability that could help build broader trust in the technology. There are currently many open questions about AI auditing: who should set the AI auditing framework? To achieve what outcomes? Should it focus on the model or the data (or both)? What role would it have in regulation, including sector-specific supervision? There is also the challenge of aligning AI auditing models with other applicable standards – such as sectoral guidance, data protection, equality law, employment law, competition policy, and international alignment of approaches.

154. There may also be a case for data auditing, including provenance, ageing, how data are generated, what changes were made to the lifecycle and what the known limitations are. There remains a skills gap for auditors when it comes to AI models and data. This is compounded by the fact that AI auditing is in its infancy and, in the immediate future, auditing would need to be carried out by multidisciplinary teams. There is also a need to consider data and model auditing holistically.

155. There are similar questions when considering AI certification: whom would the certification be aimed at and for what purpose? Who undertakes the certification and what is the role of the public and private sectors? What should the standards be and how do standards evolve over time? How to navigate trade-offs, such as the need for transparency and prevention of the loss of intellectual property? There are also challenges in fitting any new AI or data auditing and certification frameworks into an organisation's existing Enterprise-Wide Risk Management framework.

156. To ensure an effective AI audit, a framework must be in place around which assurance products could be developed. There are certainly lessons to be learned from financial audit and control frameworks, including good practice for controls and the use of data and AI.

157. More fundamentally, developing an audit framework means establishing standards to audit against first. AI audit currently has no industry standards on required metrics and, more broadly, on transparency and explainability. There is an understanding of what relevant risk frameworks should look like but they are not in place yet. There is now an opportunity to simplify metrics and approaches. For example, for an accountable executive looking after a suite of models and faced with a range of metrics, being able to interpret them becomes very difficult. As an industry, it would be useful to pick a small set of representative metrics to become commonplace. Additionally, a two-tiered approach to such standard-setting could couple high-level principles from regulators or other standard-setting bodies with more detailed internal standards embedded within firm governance processes.

158. As with other aspects of AI governance, internal audit and 3LOD teams may not have the relevant AI skills. So there may be a need for external audit conducted by third-parties with the relevant skills and this will likely increase in the coming years. External audits could be time and resource intensive but they are important, both for auditing the models themselves and the MRM processes, controls and functions.

159. The ICO has published an AI auditing framework,[67] the CDEI has published a guide to AI assurance,[68] and most major consultancies have AI audit offerings. The financial services industry in general is gearing up for AI auditing, spurred on by the need for improved risk management. A good initial step would be to gather examples of best practice together to show 'what good looks like'. This might be a more useful approach than creating a complex set of principles. Similarly, standard-setting bodies have an important role to play when it comes to effective AI auditing and regulation.

(67) AI Auditing Framework (ICO).
(68) AI assurance guide (cdeiuk.github.io).

Suggestions for good practice – Governance

Firms should:

- Strengthen contact between data science teams and risk teams from the early stages of the model development cycle while maintaining independent challenge.
- Establish a central committee to oversee firm-wide development and use of AI. Firms should have AI-specific elements of the risk framework; as wells as a privacy framework; operational principles; and, ideally, a set of ethical principles which can help guide decision-making.
- Provide training and understanding of AI (including responsibility and accountability) with the aim of embedding a sufficient level of skills throughout the organisation.
- Should share good practice across the organisation, for example, from data management, through to software development and MLOps.

Conclusion and next steps

160. The AIPPF discussions on the nature and uses of AI have been broad and deep, mirroring wider debates taking place across the financial services sector and beyond. The AIPPF meetings, workshops, and ad-hoc discussions have highlighted the benefits as well as the many complex challenges in adopting and using AI. The Forum also brought together diverse views on potential ways of addressing those challenges.

161. While this report has focused largely on the role of Data, Model risk, and Governance in the adoption and use of AI in financial services, these sit within domestic and international regulatory and legislative frameworks. Clarity of regulatory expectations on the adoption and use of AI is a key component of fostering innovation. Regulators should provide greater clarity on existing regulation and policy. Such clarification and any new guidance should not be overly prescriptive and should provide illustrative case studies. Alongside that, regulators should identify the most important and/or high-risk AI use-cases in financial services with the aim of developing mitigation strategies and/or policy initiatives.

162. In terms of next steps, it is clear that AI will continue to develop rapidly. Regulators and industry practitioners should continue to monitor and support the safe adoption of AI in financial services. Public-private engagement is invaluable and should continue with a wide range of stakeholders, including representation from civil society through regular or ad hoc forums. It would also be useful to have more structured and regular engagement on best practice or industry guidelines with a formal consultation process allowing for feedback.

163. An industry consortium could serve as a next step towards developing industry solutions to specific challenges and to creating industry-wide standards. Establishing an organisation to certify AI practitioners may also be useful and complementary to algorithm certification/auditing.

Annexes

I – Use-cases

164. It may be instructive to consider the various issues highlighted in this paper in the context of three specific use-cases in financial services: (i) credit, (ii) savings and investment advice, and (iii) AML and fraud detection. Other use-cases worth considering include: customer on-boarding; money management and personalised financial offers; information extraction and document scanning; and marketing.

165. While considering existing use-cases is useful, it should also be noted that one of the reasons AI is important is that it can enable new use-cases. For example, high-dimensional and quantum-like algorithms are starting to appear (though not necessarily in financial services). These could be applied to complex money-laundering networks, especially where parts of the network are hidden.

166. There are various benefits to the use of AI in each of the three use-cases:

166.1. On credit, AI models can potentially help improve predictive power and firm profitability. AI credit-card models can result in an uplift in predictive power of up to 10% (as measured by Gini impurity)(69) compared to logistic regression models.(70) The benefit to the firm is additional revenue and the benefit to the consumer is lower false positives/negatives.

166.2. On savings and investment advice, AI models can potentially provide much more granular and personalised advice. They can also help consumers better understand the potential risks and make more informed investment choices with appropriate guardrails.

166.3. On AML and fraud detection, AI can help address the challenge of synthetic identity fraud, whereby synthetic identities are created from a jigsaw of real data.

167. The following table provides more detail on the main types of AI models and data used in the three use-cases:

Use-case	Data	Models
Credit	Traditional credit applications, complimented with alternative data (telecoms, social media, etc.). Behavioural data.	Vast majority are linear regression. However, with AI/ML they tend to be gradient boosted models, typically decision trees. Neural networks which may be used as inputs to linear regression models.
Savings & investment advice	Lots of customer data, market signal data and alternative data.	Natural language processing is used for sentiment analysis.
AML & fraud detection	Very broad data sources, everything from IP address to behavioural patterns.	AML and sanctions use automated tree-based models, such as random forest, and graph-based AI. Fraud detection use neural networks and lots of unsupervised models to generate alerts.

(69) Gini impurity (not to be confused with the Gini coefficient in economics) is a measure of a model's power to classify new instances.
(70) Logistic regression is a statistical technique that uses a logistic function to model a binary dependent variable.

168. It is worth noting that in practice, retail credit decisioning is not yet fully automated. AI-based models support decisions that act as inputs into the overarching credit lending process. This is partially due to the treating consumers fairly principle and the amount of human control and oversight currently required to meet the principle.

169. There is an expectation that AI models would be trained on domestic data (even though not specifically required by domestic regulators) in the absence of global rules.

170. The majority of challenges and risks apply to all three use-cases and so the following comments should be viewed as common issues for managing and mitigating AI risks, with the exception of:

170.1. For credit, the use of AI may contradict certain Basel requirements. In particular, firms are required to use the same logic (and models) for external credit assessments and internal ratings based (IRB) capital calculations (the 'use test'). Some banks have moved to AI models for the former but not the latter, which may need addressing within the regulatory framework.

170.2. For savings and investment advice, many firms have frameworks to ensure fiduciary duty and mitigate against misselling. Given the potential risks to consumers and firms, there could be merit in combining fiduciary aspects with MRM when AI models are used for retail financial services, such as recommending investment products.

170.3. For AML and fraud detection, a fundamental problem is that firms do not have access to 'ground truths' since regulators do not typically share which Suspicious Activity Reports were genuine and which were not. This lack of relevant information impacts the training as well as overall performance of the model and its wider applicability.

171. The table below summarises the key risks to consumers, to firms, and to the financial system, for each of the use-cases.

Use-case	Risk to consumers	Risk to firms	Systemic risk
Credit	Mistakes by the model, which lead to products and services being wrongly denied (false positives). Mistakes, which lead to higher-risk customers being given access to credit (false negatives).	Liability and/or reputational risk for denying credit to credit-worthy customers.	Lack of diverse credit data (in UK), which could lead to systemic credit risk. (e.g. Covid payment deferrals are hidden in credit bureau data but could cause issues).
Savings & investment advice	Wrong advice, which leads to poor performing investments and, potentially, higher losses. Use of personal data that could lead to discrimination and refusal of services.	Liability and/or reputational risk for wrong advice. Personal data used inappropriately, which could breach GDPR and lead to fines.	Herd behaviour.
AML & fraud detection	Mistakes, which lead to payments or transactions being wrongly denied (false positives). Mistakes, which lead to payments or transactions being wrongly granted (false negatives).	Liability for denying payments or services to legitimate customers. Reputational and potential conduct risk.	

172. For all three use-cases, AI changes the speed and scale at which decisions can be made – with both potential positive as well as negative effects. For example, a human investment manager could only give inadequate investment advice to a small number of customers over a certain period of time. However, an AI investment model could potentially provide poor advice to a significant number of retail consumers in a short period of time. On the systemic level, this could potentially lead to herd behaviour and/or artificially inflate or deflate the value of certain assets while the faulty algorithms are in operation.

173. The complexity of existing IT architectures applies to all of the use-cases. Moreover, this means that errors and risks (such as operational failures) can occur when deploying AI models into production. It also makes it more difficult to identify errors after an AI model has been deployed, and can take longer to find and fix. This is particularly true for larger firms with legacy systems.

174. The interconnected nature of AI models, where the outputs of one model become the input of another, means there is also a risk that the reward scenario for one model could drive unwanted behaviour in another model.

175. Both data drift and concept drift could impact all three use-cases and potentially cause risks to firms and/or the wider financial system.

176. The following risks to consumers could also apply to all use-cases:

176.1. Financial exclusion: AI systems may prevent certain customers from accessing a financial product or service. They may restrict customers' ability to get credit or insurance cover; their ability to access certain investment products or even their ability to enter into a relationship with the financial institution. AI systems may also prevent customers from enjoying one or more benefits that they can reasonably expect from an existing product or relationship such as their claims against an insurance policy, their ability to make payments or other transactions.

176.2. Competition concerns: consumers may experience unfavourable commercial outcomes compared to others when applying for or using a product or service. This could impact pricing, penalties, product conditions (such as tenure), or level of collateral.

176.3. Fiduciary duty not met: AI systems may increase the risk of breaching a financial institution's fiduciary duty to consumers, such as the fair treatment principle,(71) affordability, treatment of vulnerable consumers, as well as fair complaints processing.

176.4. Breaching customers' personal data rights: AI systems may lead to incremental disclosure of protected data or inappropriate engagement with customers that go against previously agreed terms and conditions (e.g. when a customer's video call is used by an algorithm to detect emotions without explicit consent).

177. Furthermore, the following risks to firms could apply to all use-cases:

177.1. Financial loss (e.g. from a poor credit algorithm).

177.2. Reputational risk (including perceived reputation risk e.g. Apple Card).(72)

177.3. Regulatory risk (risk of being fined or sanctioned because of a breach of responsibilities to customers).

177.4. Resilience risk (being unable to recover quickly from material disruption).

177.5. Risk of loss of intellectual property.

(71) Fair treatment of customers (FCA).
(72) Report on Apple Card Investigation (New York State Department of Financial Services, 2021).

II – BCBS 239 Principles(73)

Principle	Summary
I – Overarching governance and infrastructure	
1. Governance	Risk data aggregation capabilities and risk reporting practices should be subject to strong governance arrangements.
2. Data architecture and IT infrastructure	Design, build, and maintain data architecture and IT infrastructure which fully supports risk data aggregation capabilities and risk reporting practices not only in normal times but also during times of stress or crisis.
II – Risk data aggregation capabilities	
3. Accuracy and integrity	Generate accurate and reliable risk data to meet normal and stress/crisis reporting accuracy requirements. Data should be aggregated on a largely automated basis so as to minimise the probability of errors.
4. Completeness	Capture and aggregate all material risk data across the group. Data should be available by business line, legal entity, asset type, industry, region, and other groupings, as relevant for the risk in question, that permit identifying and reporting risk exposures, concentrations and emerging risks.
5. Timeliness	Generate aggregate and up-to-date risk data in a timely manner while also meeting the principles relating to accuracy and integrity, completeness and adaptability.
6. Adaptability	Generate aggregate risk data to meet a broad range of on-demand, ad hoc risk management reporting requests, including requests during stress/crisis situations, requests due to changing internal needs and requests to meet supervisory queries.
III – Risk reporting practices	
7. Accuracy	Risk management reports should accurately and precisely convey aggregated risk data and reflect risk in an exact manner. Reports should be reconciled and validated.
8. Comprehensiveness	Risk management reports should cover all material risk areas within the organisation. The depth and scope of these reports should be consistent with the size and complexity of the operations and risk profile, as well as the requirements of the recipients.
9. Clarity and usefulness	Risk management reports should communicate information in a clear and concise manner. Reports should be easy to understand yet comprehensive enough to facilitate informed decision-making. Reports should include meaningful information tailored to the needs of the recipients.
10. Frequency	The Board and Senior Management should set the frequency of risk management report production and distribution. Frequency should reflect the needs of the recipients, the nature of the risk, and the speed, at which the risk can change, as well as the importance of reports to sound risk management and effective and efficient decision-making.
11. Distribution	Risk management reports should be distributed to the relevant parties while ensuring confidentiality is maintained.

(73) Principles for effective risk data aggregation and risk reporting (BCBS).

III – List of AIPPF members

Name	Organisation
Michael Baldwin	Ex-Google
Jason Barto	Amazon Web Services
Fiona Browne	Datactics
Javier Campos	Experian
Hugh Christensen	Amazon Web Services
Cosmina Dorobantu	The Alan Turing Institute
Mike Dewar	Mastercard
Sarah Gadd	Credit Suisse
Dan Kellett	Capital One UK
Rachel Kirkham	MindBridge AI
Shameek Kundu	Truera
Jessica Lennard	Visa
Andy Moniz	Acadian Asset Management
Owen Morris	Aviva
Gwilym Morrison	Royal London
Harriet Rees	Starling Bank
Kate Rosenshine	Microsoft UK
Jas Sandhu	Royal Bank of Canada
Amy Shi-Nash	National Australia Bank
Phil Tetlow	IBM UK
Philip Treleaven	University College London

Glossary and acronyms

Glossary

This glossary should not be considered an indication of regulatory definitions. The definitions and explanations contained herein are only to clarify references to the associated concepts in the report.

AI system	A collection of AI algorithms, models, and/or applications designed to work in a coordinated way towards specific objectives.
Algorithm	A finite sequence of well-defined instructions.
Alternative and unstructured data	Data, usually unstructured and non-financial data, not traditionally used in financial modelling, including satellite imagery, telemetric or biometric data, and social-media feeds. These data are unstructured in the sense that they do not have a defined data model or pre-existing organisation.
Application	A computer programme designed to carry out specific tasks.
Explainability	The idea that an AI model or system can be explained in a way that is meaningful to a human being.
Features	Characteristics, attributes, or properties extracted from raw data that are then used as model inputs in singular or aggregated form.
Federated learning	A method of training models across distributed and/or decentralised networks using local data and without exchanging those data. Effectively, models are transferred rather than data thereby masking individual data records.
Herding	Tendency of market participants to follow and imitate the behaviour of other market participants.
Hyperparameter	Parameters used in AI and machine learning to control the model selection and training process. They are specified prior to training iterations unlike parameter values calculated as part of model fitting.
LIME	Local Interpretable Model-agnostic Explanations is a technique for approximating a machine learning model with a local, interpretable model to help explain individual outputs.
MLOps	A set of processes at the intersection of machine learning, data engineering and DevOps (the combination of software development and operational processes to provide efficient, continuous software delivery).
Model	A quantitative method, system, or approach that applies statistical, economic, financial, or mathematical techniques to process input data into quantitative or qualitative outputs.
Model risk	Potential loss a firm may incur as a result of decisions principally based on the outputs of models; this may be due to errors in the specification, development, implementation, or use of such models.
Procyclical behaviour	The tendency for market participants to act in a way that reinforces and magnifies market or financial fluctuations.
Reinforcement learning	One of the three classic machine-learning paradigms, in which an agent within the model takes actions based on the state of its environment to optimise a utility function.
Reproducibility	A key tenet of the scientific method, this is the principle that results and findings should be documented in such a way that calculations can be repeated to produce statistically identical results.
SHAP	A Shapley Additive Explanation is a tool used to quantify and help explain or interpret the influence particular variables have on a model's outputs.
Supervised learning	One of the three classic machine-learning paradigms, in which the model maps input variables to output variables by learning from labelled training samples. Example models include Support Vector Machines and Neural Networks.
Synthetic data	Data generated algorithmically rather than from actual events or direct measurement.
Three lines of defence	The lines of defence are the governance and controls to protect against risks in an organisation. The First Line of Defence is risk mitigation and control within the business function that generates the risks. The Second Line of Defence is an independent oversight function – commonly the risk functions monitoring each key risk category. The Third Line of Defence is an independent assurance function – the internal audit function.
Unsupervised learning	One of the three classic machine-learning paradigms, in which the model aims to identify patterns in the data. Examples include Clustering and Principal Component Analysis.

Acronyms

1LOD	First-Line-of-Defence
2LOD	Second-Line-of-Defence
3LOD	Third-Line-of-Defence
AML	Anti-Money Laundering
BCBS	Basel Committee on Banking Supervision
CDEI	Centre for Data Ethics and Innovation
D-SIB	Domestically Systemically Important Bank
EC	European Commission
FMSB	Fixed Income, Currencies and Commodities Markets Standards Board
G-SIB	Globally Systemically Important Bank
HMT	Her Majesty's Treasury
ICO	Information Commissioner's Office
ISO	International Organization for Standardization
IRB	Internal Ratings-Based (modelling approach)
LIME	Local Interpretable Model-agnostic Explanation
MIFID II	Markets in Financial Instruments Directive 2
MLRO	Money Laundering Reporting Officer
MRM	Model Risk Management
OAI	Office for AI
PRA	Prudential Regulation Authority
SHAP	Shapley Additive Explanation
SM&CR	Senior Managers & Certification Regime