Financial Stability Institute

FSI Insights
on policy implementation
No 35

Humans keeping AI in check – emerging regulatory expectations in the financial sector

By Jermy Prenio and Jeffery Yong

August 2021

JEL classification: C60, G29, G38, O30
Keywords: artificial intelligence, machine learning, corporate governance, risk management, risk modelling

BANK FOR INTERNATIONAL SETTLEMENTS

FSI Insights are written by members of the Financial Stability Institute (FSI) of the Bank for International Settlements (BIS), often in collaboration with staff from supervisory agencies and central banks. The papers aim to contribute to international discussions on a range of contemporary regulatory and supervisory policy issues and implementation challenges faced by financial sector authorities. The views expressed in them are solely those of the authors and do not necessarily reflect those of the BIS or the Basel-based committees.

Authorised by the Chair of the FSI, Fernando Restoy.

This publication is available on the BIS website (www.bis.org). To contact the BIS Media and Public Relations team, please email press@bis.org. You can sign up for email alerts at www.bis.org/emailalerts.htm.

ISSN 2522-2481 (print)
ISBN 978-92-9259-496-1 (print)

ISSN 2522-249X (online)
ISBN 978-92-9259-497-8 (online)

Contents

Executive summary 1

Section 1 – Introduction 3

Section 2 – Authorities' expectations or guidance relating to the use of AI by financial institutions 6

Section 3 – Are AI-related expectations or guidance captured in existing standards or regulations? 10

Section 4 – Challenges in implementing the AI-related expectations or guidance 14

Transparency 14

Reliability and soundness 15

Accountability 16

Fairness and ethics 17

Addressing regulatory/supervisory challenges through proportionality 18

Section 5 – Conclusion 20

Annex – Proposed AI regulation in the EU 22

References 24

Humans keeping AI in check – emerging regulatory expectations in the financial sector[1]

Executive summary

Artificial intelligence (AI) including machine learning (ML) can offer significant potential to improve the delivery of financial services and operational and risk management processes. Technology is now part and parcel of financial services and there is no question that it will continue to drive profound changes for consumers and financial institutions. This trend has been supported by financial authorities' efforts to promote innovation and use of new technologies in the financial sector. In doing so, sound regulatory frameworks are essential to optimise benefits and minimise risks from these new technologies.

There are AI governance frameworks or principles that apply across industries and, more recently, several financial authorities have initiated development of similar frameworks for the financial sector. Within these frameworks, several common themes converge on general guiding principles on reliability, accountability, transparency, fairness and ethics. Other guiding principles mentioned in certain frameworks relate to data privacy, third-party dependency and operational resilience. While such high-level principles are useful in providing a broad indication of what firms should consider when using AI technologies, there are growing calls for financial regulators to provide more concrete practical guidance. One approach to meet this industry need is for regulators to provide compilation of emerging industry best practices on AI governance, where available, for each of these generally accepted principles.

Existing requirements on governance, risk management, as well as development and operation of traditional models also apply to AI models. These include governance requirements placing responsibility on the use of such models on boards and senior management of financial institutions. More specifically, firms are typically required to have sound model validation processes in place to ascertain the reliability of modelling results. Importantly, supervisors expect such models to be transparent, not only as part of good risk management practice, but also to enable supervisory review of models. Moreover, existing laws, standards or regulatory guidance cover data privacy, third-party dependency and operational resilience, including in the use of models.

While most of the issues arising from the use of AI by financial institutions are similar to those for traditional models, the perspective might be different. In the context of AI models, some of the common guiding principles identified above are viewed from the perspective of fairness. For example, ensuring reliability/soundness of AI models aims to avoid causing discrimination due to inaccurate decisions. Moreover, ensuring accountability and transparency in the use of AI includes ascertaining that data subjects are aware of data-driven decisions, and have channels to inquire about and challenge these decisions.

The stronger emphasis on fairness in the use of AI results in calls for more human intervention. The general thrust of AI pronouncements by regulatory bodies seems to focus on undesirable results, including unintended bias that lead to discriminatory outcomes that can arise from automations and lack of transparency in AI models. Although human-based modelling can suffer from similar weaknesses of AI models – for example bias or errors – a key distinguishing feature of the former is that humans can be held unambiguously accountable. Use of AI, however, can lead to philosophical

[1] Jermy Prenio (Jermy.Prenio@bis.org) and Jeffery Yong (Jeffery.Yong@bis.org), Bank for International Settlements. We are grateful to Douglas Araujo, Julian Arevalo, Orlando Fernandez Ruiz, Denise Garcia Ocampo, Xuchun Li and Oliver Thew for helpful comments. Luciana D'Agnone provided valuable administrative support.

reflection of the demarcation between machines and human beings. This challenge is compounded by a number of factors, including (i) the speed and scale of AI adoption by financial institutions; (ii) technical construct of AI algorithms; and (iii) lack of model explainability. From a regulatory standpoint, placing accountability firmly and clearly on responsible persons within a firm is key to operationalising sound AI regulatory frameworks. However, trade-offs will need to be made between reaping the benefits from large scale machine automation against the need for human input and oversight.

There is scope to further define fairness to support sound AI governance. Fairness, as described in the policy documents covered in the paper, relates to avoiding discriminatory outcomes. However, non-discrimination may not be explicit in consumer protection laws in some jurisdictions. Making non-discrimination objectives explicit may help provide a good foundation for defining fairness in the context of AI, provide a legal basis for financial authorities to issue AI-related guidance and, at the same time, ensure that AI-driven, traditional model-driven and human-driven decisions in financial services are assessed against the same standard.

The challenges and complexity presented by AI call for a proportional and coordinated regulatory and supervisory response. This requires differentiating the regulatory and supervisory treatment on the use of AI models, depending on the conduct and prudential risks that they pose. AI models whose results have significant implications on conduct and prudential risks will need to be subject to more stringent regulatory and supervisory treatment than those with less significant implications. In addition, use of AI by financial institutions will have implications for profitability, market impact, consumer protection and reputation. This calls for more coordination between prudential and conduct authorities in overseeing the deployment of AI in financial services.

Given emerging common themes on AI governance in the financial sector, there seems to be scope for financial standard-setting bodies to develop international guidance or standards in this area. Authorities' views on how these common themes should be implemented are still evolving. A continued exchange of views and experiences at the international level could eventually lead to the development of international standards. Such international standards could be helpful particularly to jurisdictions that are just starting their digital transformation journey. They can also serve as a minimum benchmark in guiding orderly deployment of AI technologies within the financial sector. As more specific regulatory approaches or supervisory expectations emerge on specific aspects of AI use cases, the standard-setting bodies can identify such common "best practice" that will be useful for other jurisdictions to consider. At the same time, given the evolving technology trends, principles-based guidance continue to have its benefits and can complement such best practice approach.

Section 1 – Introduction

1. **Artificial intelligence (AI) including machine learning (ML) is one of the defining technologies that is reshaping the financial sector.** Within regulatory circles, a commonly quoted definition of AI is by the Financial Stability Board (2017), which states "the application of computational tools to address tasks traditionally requiring human sophistication is broadly termed 'artificial intelligence'". The paper further described that ML can be defined as a "method of designing a sequence of actions to solve a problem, known as algorithms, which optimise automatically through experience and with limited or no human intervention." Much like how the internet transformed the way we bank or shop for insurance, AI has the potential to significantly improve delivery of financial services, but it also brings new risks that financial sector supervisors must grapple with.[2]

2. **AI technology can significantly improve the delivery of financial services to consumers as well as the operational and risk management processes within firms.**[3] Examples of AI use cases in terms of consumer benefits include:

- widening access to credit
- robo-advisers that provide automated investment advice based on a consumer's investment goals and risk profiles
- chatbots that provide instant response to basic customer queries
- more efficient insurance claims processing

Internally within firms, AI technology promises huge potential in the following areas:

- identification of suspicious financial transactions that could be fraudulent, money laundering or terrorism financing
- risk scoring that enables automated loan granting (in some cases even without collateral) or insurance pricing decision
- risk management and/or calculation of regulatory capital requirements

3. **Although AI can benefit both consumers and firms, it can also introduce and/or exacerbate risk exposures.** Risks such as unintended bias or discrimination against certain groups of consumers are not only a market conduct or consumer protection issue, but they also concern prudential supervisors when such risks translate into financial exposures for firms or if they give rise to large-scale operational risks, including cyber risk and reputational risk. In addition, prudential risks can arise from wide-scale under-pricing of financial products or systematic errors in underwriting new financial consumers. Ultimately, safeguards must be in place to protect consumers' interests and to maintain the safety and soundness of financial institutions.

4. **As more financial institutions are increasing the use of AI to support their business processes, financial regulators are starting to put in place or update specific regulatory frameworks on AI governance.** With a few exceptions like in the European Union (EU) where there is already a legislative proposal to harmonise the rules for AI, most frameworks are still in early stages of development, and range from application of existing principles-based corporate governance requirements in an AI context to practical non-binding supervisory guides on how to manage AI governance risks. In several cases, these frameworks complement or intersect with cross-sectoral frameworks on AI governance developed by non-financial regulators such as cybersecurity and/or data protection authorities. While some may argue that, from a regulatory standpoint, the use of AI is nothing new (for example, it is

[2] For a micro- and macro-analysis of potential effects of the adoption of AI in financial markets, see FSB (2017).

[3] See for example EIU (2020).

comparable to the adoption of internal models by firms), the scale and speed of AI adoption warrants special regulatory attention. Having sound AI governance frameworks is increasingly important given progressively wider adoption by firms and the growing number of financial authorities promoting innovation and technology in the financial sector. Financial innovation should not compromise the core mandates of financial sector supervisors.

5. **This paper canvasses a selection of policy documents (Table 1) on AI governance issued by financial authorities or groups formed by them in nine jurisdictions and other cross-industry AI governance guidance that apply to the financial sector.** The paper aims to provide a snapshot of existing regulatory approaches on AI governance, including initial thinking as expressed in consultation papers and to identify emerging common regulatory themes including from relevant cross-industry, general AI guidance. The policy benchmarking exercise is supplemented by interviews with four financial authorities.[4] Other regulatory authorities may draw inspiration from this stock take to consider adopting relevant regulatory approaches that are appropriate for their local circumstances.

[4] The European Banking Authority, the European Insurance and Occupational Pensions Authority, the Monetary Authority of Singapore and the UK Prudential Regulation Authority.

AI-related issuances applicable to the financial sector

Table 1

	Regulation/legislation	Guidance; guidelines	Principles	Discussion paper; others
European Union	✓(EC[1])	✓(HLEG[2])	✓(EIOPA[3])	✓(EBA[4], EIOPA[5])
France				✓(ACPR[6])
Germany			✓(BaFin[7])	✓(BaFin[8])
Hong Kong, SAR			✓(HKMA[9])	
Luxembourg				✓(CSSF[10])
Netherlands			✓(DNB[11])	
Singapore			✓(MAS[12])	
United Kingdom		✓(ICO[13])		✓(BoE/FCA[14])
United States			✓(NAIC[15])	✓(UST[16], US Agencies[17])
International			✓(OECD[18], G20[19])	

[1] European Commission, Proposal for a regulation laying down harmonised rules on AI (April 2021).

[2] Independent High-level Expert Group on AI (set up by the European Commission), Ethics guidelines for trustworthy AI (April 2019).

[3] European Insurance and Occupational Pensions Authority, Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the European insurance sector (June 2021).

[4] European Banking Authority, Report on big data and advanced analytics (January 2020).

[5] European Insurance and Occupational Pensions Authority, Big data analytics in motor and health insurance: A thematic review (May 2019).

[6] French Prudential Supervision and Resolution Authority (ACPR), Governance of AI in Finance (June 2020).

[7] Federal Financial Supervisory Authority of Germany (BaFin), Big data and artificial intelligence: Principles for the use of algorithms in decision-making processes (June 2021).

[8] Federal Financial Supervisory Authority of Germany (BaFin), Big data meets AI (July 2018).

[9] Hong Kong Monetary Authority, High-level principles on AI (November 2019); Consumer protection in respect of Use of Big Data Analytics and Artificial Intelligence by Authorized Institutions (November 2019).

[10] Financial Sector Supervisory Commission of Luxembourg (CSSF), AI: Opportunities, risks and recommendations for the financial sector (December 2018).

[11] Netherlands Bank, General principles for the use of AI in the financial sector (July 2019).

[12] Monetary Authority of Singapore, Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of AI and data analytics in Singapore's financial sector (November 2018).

[13] UK's Information Commissioner's Office, draft Guidance on the AI auditing framework (February 2020) and Guidance on AI and data protection (July 2020).

[14] Bank of England and Financial Conduct Authority, Machine Learning in UK financial services (October 2019).

[15] National Association of Insurance Commissioners (2020), Principles on Artificial Intelligence.

[16] US Treasury, A financial system that creates economic opportunities: nonbank financials, fintech, and innovation (July 2018).

[17] US regulatory agencies, Request for information and comment on financial institutions' use of AI, including machine learning (March 2021).

[18] Organisation for Economic Cooperation and Development, AI Principles (May 2019).

[19] G20, AI Principles (June 2019).

6. **The rest of the paper is structured as follows.** Section 2 summarises selected authorities' expectations on AI governance, while Section 3 highlights relevance of existing international standards in this context. Section 4 describes challenges in the implementation of sound governance frameworks surrounding the use of AI, and Section 5 concludes with key takeaways.

Section 2 – Authorities' expectations or guidance relating to the use of AI by financial institutions

7. **As shown in Table 1, most of the issuances covered in this paper are in the form of discussion papers and principles.** At the international level, one of the main recommendations and principles that are commonly referenced in the financial sector are the Organisation for Economic Cooperation and Development (OECD) AI Principles[5] and the G20 AI Principles[6] – the latter drawing from the former's recommendations. These principles cover five broad areas, namely: benefits to people and planet; respect for rule of law and human rights; transparency and responsible disclosure; continuous risk assessment and accountability. In essence, these principles call for the use of AI that truly benefits society at large. AI-related guidance have also been issued at the regional or national levels (ie the EU and the United Kingdom respectively). Principles that specifically target the use of AI in the financial sector have been issued in the European Union, Germany, Hong Kong, the Netherlands, Singapore and United States. More recently, a proposed regulation laying down harmonised rules on AI across all industries, which is a first in the world, was issued in the EU (see Annex).

8. **The existing issuances revolve around five common principles – reliability/soundness, accountability, transparency, fairness and ethics**.[7][8] Table 2 summarises the regulatory expectations relating to these common principles. Other issues covered by the issuances include data privacy, third-party dependency and operational resilience. Most of these issues are the same ones authorities look at when assessing traditional models used by financial institutions.[9] What follows is a discussion of these issues and how authorities may be approaching them differently in the case of AI vis-à-vis traditional models.

[5] See OECD (2019).

[6] See G20 (2019).

[7] It is acknowledged that the different authorities may use other terms to characterise similar concepts or may group certain concepts together (eg reliability/soundness under fairness). How the paper calls or distinguishes the different concepts are based on the authors' judgment.

[8] See Fjeld et al (2020) for a more extensive review of AI-related principles, including by the private and non-financial sectors.

[9] For purposes of this paper, "traditional models" refer to models used by financial institutions that do not use AI or ML algorithms and that authorities have examined or issued guidelines for. These include, for example, models used for risk assessment or regulatory capital calculations.

Summary of regulatory expectations relating to the AI common principles Table 2

Common principles	Regulation/legislation/guidance
Reliability/soundness	• Similar expectations as those for traditional models (eg model validation, defining metrics of accuracy, updating/retraining of models, ascertaining quality of data inputs) • For AI models, assessing reliability/soundness of model outcomes is viewed from the perspective of avoiding causing harm (eg discrimination) to consumers
Accountability	• Similar expectations as outlined in general accountability or governance requirements, but human involvement is viewed more as a necessity • For AI models, accountability includes "external accountability" to ascertain that data subjects (ie prospective or existing customers) are aware of AI-driven decisions and have channels for recourse
Transparency	• Similar expectations as those for traditional models, particularly as they relate to explainability and auditability • For AI models, external disclosure (eg data used to make AI-driven decisions and how the data affects the decision) to data subjects is also expected
Fairness	• Stronger emphasis in AI models (although covered in existing regulatory standards, fairness expectations are not typically applied explicitly to traditional models) • Expectations on fairness relate to addressing or preventing biases in AI models that could lead to discriminatory outcomes, but otherwise "fairness" is not typically defined
Ethics	• Stronger emphasis in AI models (although covered in existing regulatory standards, ethics expectations are not typically applied explicitly to traditional models) • Ethics expectations are broader than "fairness" and relate to ascertaining that customers will not be exploited or harmed, either through bias, discrimination or other causes (eg AI using illegally obtained information)

Source: FSI analysis.

9. **Assessing reliability/soundness of AI models is similar to traditional models, but with emphasis on avoiding harm or discrimination.** The assessment of reliability/soundness of AI and traditional models involves looking at similar aspects such as undertaking model validation, defining metrics for accuracy, updating (or re-training, in the case of AI) of models, ensuring quality of data used by models, etc. What seems to be the distinguishing factor is that ensuring reliability/soundness of AI models is viewed from the perspective of avoiding causing harm (eg discrimination) due to inaccurate decisions arising from inherent bias in the data inputs or modelling approach.

10. **AI-related accountability issues are quite similar to general accountability or governance issues, but human involvement seems to be viewed slightly differently.** Accountability relates to having clear roles and responsibilities, as well as laying the ultimate responsibility on the board and senior management of a financial institution. In the case of AI, there is emphasis on human intervention in model development and decision making. Concepts like "human-in-the-loop" (human intervention in the decision cycle of the AI) and "human-on-the-loop" (human intervention during the design cycle and subsequent reviews) are emphasised. The role of human judgment and human review is also recognised in existing guidelines on the use of traditional models. What seems to be different is how they are perceived. In the case of traditional models (eg the internal ratings-based approach for credit risk under the Basel Framework), human overrides to models/rating assignments need to be monitored and documented in order to track the performance of overrides separately. In the case of AI, human intervention is viewed more as a necessity in order to ensure that decisions based on AI models do not result in outcomes that are unfair or unethical.

11. **Another concept of accountability that is emphasised when it comes to AI is "external accountability".** The MAS FEAT Principles[10], for example, states that financial institutions that use AI should provide data subjects (eg prospective financial customers) with channels to inquire about, submit appeals for, and request reviews of AI-driven decisions that affect them; and take into account verified and relevant supplementary data provided by data subjects when performing reviews of AI-driven decisions. While this may be a new expectation for financial institutions, the practice is not totally new. Consumers in the United States, for example, have the legal right to ask from credit reporting agencies for their credit report, which includes all the information that goes into their credit score, and to have them fixed if there are mistakes.[11]

12. **AI-related transparency issues fall into three areas:**

Explainability[12] – making transparent how an AI algorithm arrived at a certain outcome (ie disallow "black box" excuses);

Auditability – documenting AI development, processes (including decision-making) and data sets; and

External disclosure – disclosing to the data subjects the following: (i) all use of AI-driven decisions; (ii) data used to make AI-driven decisions and how the data affects the decision; and (iii) consequences of AI-driven decisions on them. OECD (2019) and G20 (2019), in particular, noted that disclosures should be in the form of plain and easy-to-understand information on the factors and logic that served as the basis for the decision in order to enable data subjects to challenge the outcome of the AI system.

13. **Explainability and auditability expectations are generally the same for both AI and traditional models and involve internal disclosure particularly to the board and senior management so they can better understand the risks and implications of AI use.** However, external disclosure expectations seem to be specific to AI use. Although as mentioned above, these are quite similar to the legal requirements relating to credit reports in the US.

14. **Fairness is not something that is typically explicitly required in the assessment of traditional models.** However, in the insurance sector, it is common for legislation to require insurers to treat customers fairly, which in principle, extends to the use of any traditional or AI models. Fairness-related issues covered in AI issuances refer to addressing or preventing biases in AI algorithms that could lead to discriminatory outcomes. The independent high-level expert group on AI set up by the European Commission also refers to a "procedural" dimension of fairness. This is quite similar to the concept of external accountability discussed above. This involves ensuring data subjects have the ability to contest and seek effective redress against AI-based decisions made by humans operating them. This implies that the entity accountable for decision-making must be identifiable and the decision-making process explainable.

15. **The concept of "fairness", however, is somewhat nebulous.[13]** Aside from stating that "fairness" can be achieved by addressing or preventing biases in order not to lead to discriminatory outcomes, the concept is not really defined in the issuances. Some might argue that there is a difference between fairness (ie equity) and bias (ie predisposition), and while the former needs to be ensured, there are "good biases" that need to be preserved (eg rewarding careful drivers whose driving behaviours are

[10] See MAS (2018).

[11] See www.myFICO.com.

[12] There is ongoing academic discussion on the difference between "explainability" as being able to explain black boxes (ie the explanation being a separate model that is supposed to replicate most of the behaviors of a black box) and "interpretability", which means designing AI models that are inherently interpretable (see Rudin, 2019).

[13] There are more than 20 mathematical definitions of fairness (see eg Verma and Rubin (2018)).

tracked by telematics with lower premiums). As discussed below, there are existing laws that aim to ensure fairness in the provision of financial services. Such laws could be used to define the concept of fairness in the context of AI but they are not common across jurisdictions. As such, some jurisdictions promote the view that individual financial institutions should define and operationalise their own "fairness" objectives, akin to having individual risk appetites (eg fairness could be defined as compliance with corporate values and ethical standards or relevant statutory requirements, such as consumer protection).[14]

16. **Some AI issuances enumerate other ways to address fairness issues in AI models.** The measures range from very general to very concrete. Some of the more concrete supervisory expectations include, for example, requiring firms to establish an ethical code of conduct to promote non-discriminatory practices; seek diversity in the input data; careful review of training and validation data during the model training process; establish policies for the procurement and lawful processing of data, especially if not available internally; have datasets that are separate from training and validation data to specifically check for model bias; embed non-discriminatory rules into the AI model; and constantly monitor the performance of the model to identify unintentional bias or to ascertain it behaves as designed and intended. There is also ongoing active work in assessing fairness of AI models.[15]

17. **Ethical issues are broader than fairness issues.** Ethics entails ensuring that customers will not be exploited or harmed, either through biases and discrimination – as in the case of fairness – or through other causes (eg AI using illegally obtained information). Ethics is based on a society's norms or mores, which may be codified in laws, regulations, codes of conduct, etc. These include privacy and data protection, non-discrimination and equality, diversity, inclusion and social justice. Another dimension of ethics relates to the question of whether AI should be deployed at all. This is specified in the DNB (2019), which calls for the objectives, standards and requirements for adopting and applying AI to be defined in an ethical code.

18. **As AI use cases increasingly utilise data inputs from a wider variety of sources including personal data, many regulatory issuances emphasise the need to comply with data privacy and protection laws/regulations.** These include the need to ensure that AI systems guarantee data privacy and protection throughout the different stages of the process, as well as to have policies governing data access and customer consent on the use of their personal data. MAS (2018) states that use of personal attributes as input factors for AI-driven decisions should be justified. EIOPA (2021) highlights that rating factors used for pricing and underwriting in insurance should have a correlation with risk and a causal link. The CSSF (2018) discussion paper goes even further by recommending the need to challenge the use of personal data as input for AI models. The BaFin (2018) discussion paper, on the other hand, points to the importance of guaranteeing freedom of choice by providing less personal data-intensive financial products.

19. **General third-party or outsourcing expectations are also relevant when it comes to AI-related third-party dependency.** Third parties may provide firms with either the data that they use for their AI models or the models themselves. Third-party risks that are relevant in the context of AI include risk to data privacy and protection, lack of understanding of how the AI model works partly due to intellectual property constraints and dependency risk. The Prudential Regulation Authority issued Supervisory Statement SS2/21, which provided AI-related examples of third-party arrangements.[16] The issuances that cover these topics emphasise that the AI policies of financial institutions should also apply

[14] See Veritas Consortium (2020).

[15] Ibid.

[16] These include the purchase of data collated by third-party providers (data brokers), eg geospatial data or data from in-app device activity, social media, etc; and "off-the-shelf" ML models open source software and ML libraries developed by third-party providers.

to third parties; that financial institutions should have third-party management frameworks in place, including the conduct of due diligence; and that financial institutions should be aware of third-party risks. The EU proposal for regulation of AI seeks to address this issue by requiring AI service providers to design and develop AI systems in a sufficiently transparent way that enables users to interpret the system's output and use it appropriately.[17]

20. **Operational resilience issues are also relevant in the context of AI use by financial institutions.** Like traditional models, use of AI models exposes a financial institution to operational vulnerabilities. These include internal process or control breakdowns, information technology lapses, risks associated with the use of third parties, model risk and cyber risk. In terms of cyber risk, AI systems can be vulnerable to "data poisoning" attacks, which attempt to corrupt and contaminate training data to compromise the system's performance.

Section 3 – Are AI-related expectations or guidance captured in existing standards or regulations?

21. **There is currently no international regulatory standards or guidance specifically on AI for the financial sector.** Nevertheless, there are existing international standards, guidance and national laws that can be applied or used as starting points in dealing with governance issues associated with AI models (see Table 3 for a non-exhaustive list for the five common principles). This is because, as mentioned, most of the issues identified above already exist in respect to use of traditional models. This is especially true when it comes to issues such as reliability/soundness, accountability, transparency, data privacy, third-party dependency, operational resilience and to a certain extent, ethics. But this is not the case when it comes to fairness, as described in the issuances covered in this paper (ie relating to non-discrimination). While there are general guidance in insurance to ensure fair treatment of customers, this is not the case for banking. Nevertheless, fairness issues are covered in consumer protection laws in some jurisdictions.

[17] Article 13 the proposed EU AI regulation.

Existing standards/laws that may be applied in implementing the AI common principles

Table 3

Common principles	Applicable standards/laws
Reliability/soundness	• Basel Core Principles (BCP) 15, Insurance Core Principles (ICP) 16, ICP 17, Basel Committee on Banking Supervision (BCBS) Principles for effective risk data aggregation and risk reporting • Minimum requirements for the use of IRB for credit risk, IMA for market risk, stress testing, technical provisions valuation
Accountability	• BCP 14, BCP 15, ICP 7, ICP 17, BCBS Corporate governance principles for banks • Minimum requirements for the use of IRB for credit risk, IMA for market risk, AMA for operational risk, stress testing, technical provisions valuation
Transparency	• ICP 17 • Minimum requirements for the use of IRB for credit risk, IMA for market risk, stress testing, technical provisions valuation
Fairness	• ICP 19, ComFrame standard 7.2a • Consumer protection laws in some countries explicitly address fairness concerns as described in AI-related issuances (ie prevent/address discriminatory outcomes)
Ethics	• BCP 29, ICP 5, ICP 7, ICP 8, BCBS Corporate governance principles for banks, BCBS Principles for the sound management of operational risk, BCBS Principles on compliance and the compliance function in banks. FSB toolkit for firms and supervisors to mitigate misconduct risk

BCP 14 Corporate governance

BCP 15 Risk management process

BCP 29 Abuse of financial services

ICP 5 Suitability of persons

ICP 7 Corporate governance

ICP 8 Risk management and internal controls

ICP 16 Enterprise risk management for solvency purposes

ICP 17 Capital adequacy

ICP 19 Conduct of business

ComFrame standard 7.2.a: The group supervisor requires the IAIG Board to ensure that the group-wide business objectives, and strategies for achieving those objectives, take into account at least the following fair treatment of customers.

Source: FSI analysis.

22. **Existing banking and insurance international regulatory standards can be applied in the context of addressing reliability/soundness of AI models.** Both the Basel Core Principles (BCP) and the Insurance Core Principles (ICP) provide guidance on reliability/soundness of risk models or models used for regulatory capital purposes.[18] BCP 15, particularly its essential criteria 6, states that banks using risk models must comply with supervisory standards on model use, including the conduct of regular and independent validation and testing of models. ICP 16 provides guidance on the use of models for risk measurement, while ICP 17 provides guidance on the use of models for regulatory capital purposes, including the assessment of appropriateness of methodology, model inputs and assumptions. The BCBS principles for effective risk data aggregation and risk reporting also provide guidance on the accurate and reliable generation of risk data by banks.[19]

[18] See BCBS (2019) and IAIS (2019), respectively.

[19] See BCBS (2013).

23. **There are also existing standards addressing reliability/soundness in specific domains where models are used for regulatory purposes.** These domain-specific (eg internal ratings-based approach (IRB) for credit risk and internal models approach (IMA) for market risk[20]; advanced measurement approach (AMA) for operational risk[21], stress testing[22]; and technical provisions valuation[23]) requirements are quite similar. They require financial institutions to ascertain that model methodology and assumptions are conceptually sound, fit for the intended purpose and have good predictive power. They also require regular validation (eg through backtesting), challenge and review. Moreover, they require vetting of data inputs to ensure accuracy, completeness and appropriateness.

24. **There are existing corporate governance standards that define general accountability requirements for banks and insurance firms.** In particular, BCP 14 for banks and ICP 7 for insurance. The BCBS also has dedicated principles on corporate governance[24]. Much more specific to model use is BCP 15 and ICP 17. Both relate to the use of risk models and assigns ultimate responsibility to the board and senior management and for them to understand the consequences and limitations of model outputs. ICP 17 also requires insurers to have adequate governance and internal controls for internal risk models used to determine regulatory capital requirements.

25. **Domain-specific standards also require effective governance structures.** This involves specifying the roles and responsibilities for all aspects of the modelling process. In the case of IRB, banks are required to have written guidance describing how human judgement and model results are combined. Banks are also required to have procedures in place for human review of model-based rating assignments that focus on finding and limiting errors associated with known model weaknesses and to improve model performance. As mentioned above, human overrides – whether of the model output itself or of the input used – must also be identified and monitored in order to separately track performance.

26. **Transparency requirements already exists for models used for supervisory purposes.** ICP 17, for example, requires insurers to document internal models used for regulatory capital purposes. The documentation should cover model design, assumptions, compliance with regulations such as use and statistical tests[25]. Domain-specific standards in banking also requires the same. In addition, they require documentation to include policies and processes followed in using model outcomes, which will form the basis of audits. The stress testing principles also requires that findings should be communicated within and across jurisdictions.

27. **Existing laws, standards or regulatory guidance sufficiently cover data privacy, third-party dependency and operational resilience, and may be applied in the context of AI.** Most jurisdictions have data privacy and protection laws, which, as mentioned above, are already referred to in AI-related issuances. In the insurance sector, ICP 19 requires insurers and intermediaries to have policies and processes for the protection and use of information on customers. In terms of third-party dependency, the Joint Forum's guiding principles on outsourcing in financial services,[26] the BCBS's principles on

[20] See BCBS (2019).

[21] Although use of AMA will be removed as an option for calculating regulatory capital under the Basel framework by 2023.

[22] See BCBS (2018).

[23] See IAIS (2019), particularly the Common Framework for the Supervision of Internationally Active Insurance Groups (ComFrame) Standard 16.7.d.

[24] See BCBS (2015).

[25] A "use test" is the process by which an internal model is assessed in terms of its application within an insurer's risk management and governance processes. A "statistical quality test" assesses the quantitative methodology of an internal model.

[26] The Joint Forum (2005).

operational resilience and operational risk management[27] provide guidance on this issue and national regulations typically include third-party or outsourcing management guidelines. The requirements for the use of models for IRB purposes also emphasise that use of such models obtained from third parties is not a justification for exemption from documentation. However, the issue of concentration risk arising from a few third-party providers of certain services (eg cloud) remains and could also be relevant for AI.[28] Finally, aside from third-party dependency, the BCBS's principles on operational resilience cover a broad range of issues that can cause operational vulnerabilities, such as cyber risk.

28. **Existing banking and insurance standards do cover issues relating to ethics.** In particular, ethics issues are mentioned throughout the ICPs. In banking, while there is only one BCP that mentions ethics (BCP 29 on abuse of financial services), topic-specific principles issued by the BCBS deal with the issue. Ethics requirements start with the role of the board in defining and overseeing the implementation of code of ethics or code of conduct and ensuring that staff receives appropriate ethics training (ICP 7, BCBS's corporate governance principles and principles for the sound management of operational risk). ICP 5 also requires setting high internal standards of ethics and integrity in ensuring that the requirements for the board, senior management and key persons in control functions are met. ICP 8 and the BCBS principles on compliance and the compliance function in banks[29] emphasise that compliance starts at the top. This means the board is responsible for establishing standards for honesty and integrity and should lead by example. Another reference is FSB (2018), which focuses on misconduct in the financial sector particularly those related to the manipulation of wholesale markets and retail mis-selling schemes. It provides general guidance on achieving ethical behaviour by financial institutions, which would cover AI use cases.

29. **Given that most insurance supervisors have an explicit conduct mandate, insurance standards typically cover fairness issues.** ICP 19 requires insurers and intermediaries to treat customers with due care, skills and diligence, and to have policies and procedures to treat customers fairly. It also requires insurers and intermediaries to take into account the interests of different types of consumers when developing and distributing insurance products. Moreover, ComFrame standard 7.2a requires boards of internationally-active insurance groups to set business strategies and objectives that give due regard to fair treatment of customers.

30. **Banking standards focus on prudential matters and, thus, do not directly cover fairness issues.** The IRB requirement that banks should vet data inputs to assess its appropriateness for assigning ratings may be viewed or applied in the context of fairness. It could also be argued that fairness concerns can be addressed under broader governance standards (eg by ensuring a sound control environment and the presence of checks and balances that guide decision making) – but so far, these concerns have not been the focus of governance standards for banks. Apart from these connections, there are no other banking standards that could be viewed as addressing fairness.

31. **Consumer protection laws in some jurisdictions address fairness concerns, particularly those relating to discriminatory outcomes.** Consumer protection laws that directly address fairness concerns as described in the AI-related issuances (ie prevent/address biases to avoid discriminatory outcomes) can be found in jurisdictions such as the EU, South Africa and the US.[30] However, non-

[27] See BCBS (2021a) and (2021b), respectively.

[28] See FSB (2020).

[29] See BCBS (2005).

[30] The EU's Consumer Credit Directive observes the principles recognised in particular by the EU Charter of Fundamental Rights, including on non-discrimination based on any ground such as sex, race, color, ethnic or social origin, genetic features, language, religion or belief, etc. South Africa's Consumer Protection Act also prohibits discrimination as defined in its Constitution or in the Promotion of Equality and Prevention of Unfair Discrimination Act. In the US, the Fair Housing Act (FHA) prohibits discrimination in residential real estate-related transactions based on: race or color; national origin; religion; sex; familial status;

discrimination does not seem to be an explicit feature or may not be considered as a "basic area" of fair treatment provisions in other jurisdictions' consumer protection laws.[31] The 2011 G20/OECD High-Level Principles on Financial Consumer Protection does cover equitable and fair treatment of customers (Principle 3), and the subsequent effective approaches to support the implementation of this principle include a non-discrimination example.[32] But taking a very broad approach to discrimination might be difficult and the industry seem to prefer not to address the issue explicitly.[33]

Section 4 – Challenges in implementing the AI-related expectations or guidance

32. **As with any regulatory frameworks, AI governance principles, requirements or guidance need to be implemented by financial institutions to achieve the intended policy objectives.** While high-level principles and guidance on AI governance by international organisations and national authorities are welcomed, financial institutions are calling for more detailed guidance that illustrates how those principles can be implemented in practice. According to the authorities covered in this paper, there is little need for more general or high-level standards on AI governance, as the common principles as outlined in Section 2 are well-understood and accepted. The following paragraphs outline specific implementation challenges related to each of those common principles.

Transparency

33. **Supervisory expectations on transparency of models used by financial institutions and that such models should not be 'black boxes', are not new.** The main difference between AI and the more traditional models is the level of complexity and the lack of explainability of certain types of ML algorithms. Explaining complex ML algorithms in a way that can be understood by a supervisor can be a challenge. Supervisors need to upskill their staff to be conversant with ML techniques. Firms need to make more efforts to explain their ML models in an understandable way. One practical litmus test is to assess understanding of the general concept of the model by their board members – if this can be understood by non-technical board members, they will most likely be understood by supervisors as well. Some ML algorithms such as deep learning and neural networks are considered as 'black box'[34], which produce sensible results but are difficult to explain or proof unlike traditional statistical models. This is because such ML algorithms work through complex interactions between multiple variables, inferring attributes from data inputs and placing different weights on different data attributes. This also makes it difficult to disclose to data subjects in plain and simple language what data are used and how these affect the decision relevant to them.

34. **Transparency of an AI/ML algorithm is a pre-requisite to fulfilling some of the other sound AI governance principles.** If a model is not transparent, it will be difficult to assess its reliability,

and handicap. The Equal Credit Opportunity Act (ECOA) prohibits discrimination in credit transactions based on: race or color; national origin; religion; sex; marital status; age (provided the applicant has the capacity to enter into a contract); applicant's receipt of income from a public assistance program; and applicant's exercise, in good faith; of any right under the Consumer Credit Protection Act.

[31] Ardic et al (2011) focused their survey on consumer protection laws and regulations on what it considers as the four basic areas of fair treatment provisions in consumer protection laws, which are: (1) restrictions on deceptive advertising; (2) abusive collections; (3) unfair or high-pressure selling practices; and (4) breach of client confidentiality.

[32] G20 and OECD (2011) and (2019), respectively.

[33] CI (2013).

[34] See CRO Forum (2019).

performance and fairness apart from assessing the model's outcome against a specified benchmark. It will also be difficult to establish accountability if it is unclear which components of the algorithm are causing errors. CSSF (2018) explained that this is particularly important when using off-the-shelf solutions that automate decision-making because disputes may arise when wrong decisions are made.

35. **Financial institutions need to apply sound judgment when determining the appropriate level of transparency of their AI/ML models based on the target audience and materiality of the models' results.** CSSF (2018) highlighted that the more critical the use case is, the more important an algorithm should be transparent. Otherwise, the business user could refute the model's results and people accountable could lose trust in the model and refuse to take responsibility. Inadequate transparency could also erode consumer confidence or dissuade customers from using AI/ML-powered financial solutions. On the other hand, MAS (2018) pointed out that excessive transparency could create confusion or unintended opportunities for individuals including customers of financial institutions to exploit or manipulate AI/ML models EIOPA (2021) considers that insurers should strive to use explainable algorithms, in particular for high–impact AI applications even if this is at the expense of model performance. In some cases where there are no viable alternatives (eg processing of images, videos or texts), insurers can use alternative governance measures such as enhanced human oversight.

Reliability and soundness

36. **With respect to reliability and soundness of AI results, the main implementation challenge arises from the technical construct of AI algorithms.** A way to frame this issue is by using the data science pipeline. At each step of the data science pipeline, technical safeguards can be put in place to facilitate reliable and sound results. Following are examples of practical challenges that can arise at each step of the data science pipeline:

Data collection: inaccurate, incomplete or biased source data sets will lead to unreliable or biased modelling results; data source may not be generalised for the intended target consumer groups.

Data cleaning and processing: this is usually the most time consuming part of implementing an AI algorithm – insufficient resources allocated to this step or lack of rigour in cleaning the data will lead to the proverbial 'garbage-in-garbage-out' results. Bias present in source data, which may be hard to detect during data cleaning, or introduced through human intervention may be amplified by the AI algorithms.

Data analysis: it is not straigthforward to determine which data should be used as a training set (to train the algorithm) and which to use as a testing set (to verify the results).

37. **Given that ML techniques, by definition, learn from new data over time, another implementation challenge in maintaining reliable and sound results relates to regular and timely updating and reviewing the model outputs.** Bank of England and Financial Conduct Authority (2020) highlighted that the underlying data sources of an AI model or their statistical properties can change over time, for example due to changes in consumer behaviour arising from digital acceleration prompted by Covid-19. Supervised ML requires constant human feedback on the performance of the algorithm so that it can apply its trained 'knowledge' to new data and continue to produce reliable results. If the context changes, the training data set will need to be updated so that the algorithm can be retrained to produce reliable results[35]. Even unsupervised ML that learns and produces outputs by itself requires timely updates and reviews. The patterns or relationships revealed need to be assessed by a human expert to ascertain that they make sense.

[35] ACPR (2020a) proposes stability as being one of four principles to measure AI algorithms. Temporal drift arising from changes in the probability distribution of the input data could undermine the performance of an AI algorithm if it is not periodically re-trained using updated data.

38. **Existing regulatory requirements on model validation may need to be adjusted to accommodate certain ML techniques.** Specifically, significant changes to internal models used by financial institutions typically require supervisory re-approval. What constitutes 'changes' in the context of ML techniques may need to be redefined. For example, supervised ML adjusts the model as it learns from new data. However, it does not make sense to define each adjustment as being a 'change' to the model that requires supervisory approval unless the model performance or outputs change significantly.

39. **It can be challenging to strike the right balance between an algorithm's simplicity and its performance.** For example, 'precision' is one of the performance measures that evaluates the proportion of correctly identified results over all results of an algorithm. To improve precision, a firm could add more elements but potentially at the expense of making the model more complex and thus, more difficult to explain and disclose.

40. **Overly onerous regulatory requirements to 'prove' the accuracy of an ML algorithm may not be achievable.** ACPR (2020b) highlighted such a challenge in meeting the proposed most stringent explainability criterion that seeks to prove that an algorithm works correctly. In practice, this requires a line-by-line review of the source code, a comprehensive analysis of all datasets used, and an examination of the model and its parameters, which some firms view as unachievable. EIOPA (2021) specifies AI governance principles that tailors the intensity of governance measures to the impact of a given AI use case.

41. **Building and executing a ML algorithm requires significant exercise of technical judgment and expertise, each choice of which could yield vastly different results.** This underscores the importance of technical human expertise in developing, running and validating an AI model so that it produces reliable and sound results. The lack of technical expertise in firms could constrain the extent to which they can utilise AI technologies. Selecting appropriate performance measures is an example where sound judgement is needed. EIOPA (2021) provides an example in fraud detection, whereby insurers should decide if the objective is to maximise the prediction accuracy (number of fraudulent claims detected), reduce the number of false positives (legitimate claims wrongly labelled as fraudulent) or false negatives (claims labelled as legitimate which in the end are fraudulent).

42. **Increased use of AI/ML in business processes by financial institutions needs to be supported by a sound cyber resilience framework.** CSSF (2018) explains how a firm can be exposed to data poisoning attacks that involves altering training data set used to build a predictive ML algorithm in order to manipulate its results. Adversarial attack is another form of cyber-attack whereby altered data sets are provided by an adversary in order for the trained model, to misclassify the data. Such cyber breaches could lead to unreliable ML results, and potentially become systemic if left undetected or amplified through scalability of the model.

Accountability

43. **There are significant challenges in meeting the accountability criterion of AI implementation, not least because this often leads to philosophical reflection of the demarcation between machines and human beings.** In general, supervisors typically place ultimate responsibility of any decisions or actions by a financial institution on its board of directors and senior management, and rightly so. While this should similarly apply to AI implementation within a firm, accountability becomes less clear at lower levels of the hierarchy. For example, if an AI algorithm systemically denies loan applications from black applicants living within a certain postcode, should the AI developer be responsible, or the head of the credit department or indeed the Chief Executive Officer?

44. **The human-in-the-loop or human-on-the-loop safeguards could give rise to new governance-related risks.** ACPR (2020a) explains that a human operator who invalidates a machine's correctly identified results could be held liable for the error. There could also be a tendency of a human operator to accept a model's results to avoid having to justify any deviation of views. Human intervention

could also introduce bias and may complicate transparency of an algorithm, having to explain subjective human overrides. More generally, human intervention in AI model design and decisions necessitates the recruitment, training and retraining of staff with specialist expertise, as well as the upskilling of boards and senior management with explicit accountability for AI deployment. Such human resource needs could pose a significant challenge for financial institutions.

45. **AI/ML implementation often requires outsourcing by financial institutions to third party service providers, which poses unique challenges in implementing sound AI governance frameworks.** Interactions with third party service providers through licensing of AI systems and procurement of data (eg credit scores) could impose constraints on financial institutions arising from intellectual property rights imposed by such providers. Crisanto et al (2018) found that most financial regulatory authorities expect their regulated institutions to retain full responsibility and accountability of outsourced services. Such requirements should establish clarity as to the party accountable for deficient AI/ML results. Over-reliance on third party service providers could also lead to commercial capture and dependency risk particularly given increasing cross-selling of digital services by bigtechs making it difficult for a financial institution to insource the skills and expertise when needed, for example under a business continuity scenario.

Fairness and ethics

46. **A key challenge in implementing 'fairness' and 'ethical' guidance or principles is the lack of universally accepted definitions of these terms.** While it is appropriate that some supervisors have left it to firms to come up with their own definitions, some firms may find it difficult to do so. Even if they could, the definitions could fall short of general expectations by consumers. Other regulators such as EIOPA (2021) provides further guidance, for example the definition of 'fair use of data', which it defines as ensuring that the data is fit for purpose and respects the principle of human autonomy by developing AI systems that support consumers in their decision-making process. The importance of fairness and ethics should not be under-estimated, especially given recent strong social justice movements such as Black Lives Matter. A public backlash could not only inflict reputational damage on a financial institution but could also cause solvency or liquidity problems if clients switch providers en masse. For prudential supervisors, dealing with fairness and ethics is particularly challenging as these are usually not within their remit nor expertise.

47. **Regulations that require exercise of sound human judgement may make it challenging to implement AI/ML.** ACPR (2020a) cited an insurance example, referring to the IDD (Insurance Distribution Directive) 2016 European directive that requires insurance product distributors to "always act honestly, fairly and professionally in accordance with the best interests of their customers." It can be challenging to fulfil this requirement when using ML to identify future insurance needs of a consumer as consideration of contextual information will likely require human assessment.

48. **Unfair or unethical AI implementation could have profound implications on consumers, which include financial exclusion with consequential destruction of livelihoods**. CSSF (2018) outlined several sources of bias – algorithmic bias arising from choosing a wrong model or human bias that is reflected in a training data set. The paper provided a credit scoring example, whereby populations that are not well represented in the dataset are unlikely to receive a favourable credit score simply because the algorithm has learnt that in the past, such applicants were not granted many loans. This poses a commercial challenge to firms as they may inadvertently be excluding profitable businesses. BaFin (2018) highlighted how consumers might not even be aware that they are being discriminated against, resulting in denial of access to certain financial services. Such a situation can occur through AI/ML's predictive power of future lifestyle changes by inferring from changes in spending patterns, even without collecting any specific data on those changes.

49. **In general, AI governance principles or guidance on fairness and ethics require human-in-the-loop or human-on-the-loop.** In practice, this means financial institutions need to allocate sufficient human resources in any AI implementation to ascertain fair and ethical results. In a way, there is some irony in having humans as safeguards to unfairness and unethical AI results when the latter is basically just reflecting these human flaws. In any case, given that one of the main benefits of AI is to reduce the need for human intervention, it will be challenging to strike a right balance in fulfilling the human-in/on-the-loop expectations versus reaping the full benefits from technology automation.

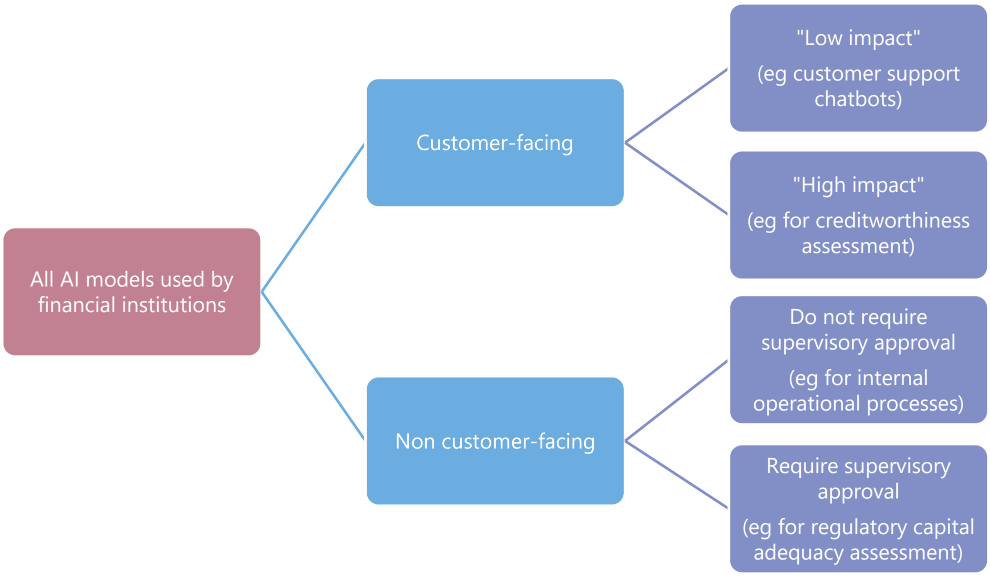
Addressing regulatory/supervisory challenges through proportionality

50. **Due to the challenges identified above, authorities are looking at applying proportionality in implementing the common principles.** This requires differentiating the regulatory and supervisory treatment of AI/ML models depending on the conduct and prudential risks that they pose. For example, AI models used for credit underwriting decisions might be subjected to higher fairness and ethical expectations than those used for customer support chatbots. Likewise, a higher level of reliability/soundness might be expected of AI models used for regulatory capital calculation relative to those used for internal operational processes. Box 1 presents a potential way forward towards implementing such a tailored regulatory and supervisory approach.

51. **There are already examples of similar tailored approach.** The proposed AI regulation in the EU identifies prohibited activities (eg AI use in the administration of justice and democratic processes) and high-risk AI systems (eg AI systems that evaluate the creditworthiness of natural persons or establish their credit score). EIOPA (2021) has developed an AI use case impact assessment framework to help insurance firms determine in a proportionate manner the combination of governance measures needed for a concrete AI use case. Effectively, higher impact activities would attract more comprehensive governance measures, and vice versa. Some authorities are also thinking of requiring board-approved overarching AI strategy if a firm's deployment of AI exceeds certain thresholds. Such strategy should be connected to relevant policies, eg codes of conduct, operational resilience, third party management, model use, etc. As such, wide deployment of AI in firms' major lines of business (eg credit or insurance underwriting) could require such a board strategy.

Box 1

Tailoring regulatory and supervisory frameworks to AI use cases


All AI models used by financial institutions
Customer-facing
Non customer-facing
"Low impact" (eg customer support chatbots)
"High impact" (eg for creditworthiness assessment)
Do not require supervisory approval (eg for internal operational processes)
Require supervisory approval (eg for regulatory capital adequacy assessment)

Regulatory and supervisory frameworks can be tailored to different AI deployment use cases, depending on whether the use case is customer-facing and materiality of its impact. The following illustrates a possible hierarchy of regulatory and supervisory frameworks to deal with different types of AI deployment by financial institutions based on these dimensions:

AI deployment	**Regulatory and supervisory treatment**
All AI models used by financial institutions***	• General governance and risk management guidelines including data privacy laws • Covered by general supervisory work of prudential supervisors
a) Customer-facing	
Low impact (eg customer support chatbots)	• In addition to regulatory and supervisory treatment in ***, requirements on fairness, ethics and data privacy apply • Covered by general supervisory work of conduct supervisors
High impact (eg creditworthiness assessment)	• In addition to regulatory and supervisory treatment in ***, more stringent requirements on reliability, soundness, accountability (including towards external stakeholders), transparency (including to data subjects), fairness, ethics and data privacy • Subject to greater scrutiny by conduct supervisors • Subject to thematic or horizontal industry reviews
b) Non customer-facing	
AI models that do not require supervisory approval (eg internal administrative processes)	• Same as ***
AI models that require supervisory approval (eg regulatory capital adequacy assessment)	• In addition to regulatory and supervisory treatment in ***, more stringent requirements on reliability, soundness, accountability (though not externally) and transparency (though not to data subjects) • Subject to greater scrutiny by prudential supervisors • Subject to thematic or horizontal industry reviews

Source: FSI staff.

Section 5 – Conclusion

52. **Most of the issues associated with the use of AI by financial institutions are quite similar to those for traditional models, but the perspective might be different.** Among the common issues identified by AI-related issuances covered in this paper, reliability/soundness, accountability, transparency, data privacy, third-party dependency and operational resilience are all relevant in the use of both AI and traditional models. Only issues related to fairness are unique or explicit to AI. However, when it comes to the use of AI, some of the common issues identified above are viewed from the perspective of fairness. For example, ensuring reliability/soundness of AI models aims to avoid causing discrimination due to inaccurate decisions. Moreover, ensuring accountability and transparency in the use of AI includes ascertaining that data subjects are aware of data-driven decisions, the data used and how it affected the decisions, and have channels to inquire about and challenge these decisions.

53. **While existing standards, laws and guidance may be used to address most AI-related issues, there may be scope to do more when it comes to fairness.** Given that most of the issues associated with the use of AI are similar to those for traditional models, authorities can leverage existing standards, laws and guidance intended for the latter in assessing the former. Fairness, particularly as it relates to avoiding discriminatory outcomes, while identified as very important in the case of AI, may not be explicit in consumer protection laws in some jurisdictions. Making non-discrimination objectives explicit may help provide a good foundation for defining fairness in the context of AI, provide a legal basis for financial authorities to issue AI-related guidance and, at the same time, ensure that AI-driven, traditional model-driven and human-driven decisions in financial services are assessed against the same standard.

54. **Most supervisors are still in early stages of developing AI-specific governance principles or guidance for financial firms.** While such high-level guidance can be useful to stand the test of time and be durable enough to apply in a technology-agnostic way (a key trait in the rapidly developing technology world), firms would find it useful to have more concrete practical guidance. Most supervisors are working in partnership with industry and other technology stakeholders to develop such guidance.

55. **In the absence of concrete practical guidance or supervisory expectations on AI governance, some firms are finding it difficult to properly set up safeguards in their AI implementations.** Challenges faced by firms are wide-ranging, from technical difficulties in building and executing ML models that deliver reliable results even under changing circumstances, to broader societal issues arising from unfair or unethical results. The first step in addressing these challenges is to strive for transparent, explainable models. This is a pre-requisite to diagnosing issues arising from reliability/soundness of results, accountability and fairness and ethics.

56. **The challenges and complexity presented by AI call for a tailored and coordinated regulatory and supervisory response based on the AI model's implications for conduct and prudential risks.** The more AI model's use can potentially impact authorities' conduct and prudential objectives, the more stringent the relevant reliability/soundness, accountability, transparency, fairness and ethics requirements should be. In addition, use of AI by financial institutions will have implications for profitability, market impact, consumer protection and reputation. This calls for more coordination between prudential and conduct authorities in overseeing the deployment of AI in financial services.

57. **Given emerging common themes on AI governance in the financial sector, there seems to be scope for financial standard-setting bodies to develop international guidance or standards in this area.** Authorities' views on how these common themes should be implemented are still evolving. A continued exchange of views and experiences at the international level could eventually lead to the development of international standards. Such international standards could be helpful particularly to jurisdictions that are just starting their digital transformation journey. They can also serve as a minimum benchmark in guiding orderly deployment of AI technologies within the financial sector. As more specific regulatory approaches or supervisory expectations emerge on specific aspects of AI use cases, the

standard-setting bodies might be in a better position to identify such common best practice that will be useful for other jurisdictions to consider.

Annex – Proposed AI regulation in the EU

The proposed regulation defines 'artificial intelligence system' (AI system) as a software that is developed with one or more of the following techniques and approaches that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with:

a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning

b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems

c) Statistical approaches, Bayesian estimation, search and optimisation methods.

The proposed regulation identifies prohibited practices, which comprise all those AI systems whose use is considered as contravening EU values. These include AI systems that manipulate persons through subliminal techniques beyond their consciousness, result in social scoring for general purposes done by public authorities, and use remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement except in certain limited situations.

The proposed regulation also identifies high-risk AI systems as those used in the following areas:

a) Biometric identification and categorisation of natural persons

b) Management and operation of critical infrastructure

c) Education and vocational training

d) Employment, workers management and access to self-employment;

e) Access to and enjoyment of essential private service and public services and benefits (included here are AI systems that evaluate the creditworthiness of natural persons or establish their credit score

f) Law enforcement

g) Migration, asylum and border control management

h) Administration of justice and democratic processes.

The proposed regulation imposes the following requirements for high-risk AI systems:

a) Risk management system – for credit institutions regulated under the Capital Requirements Directive, the risk management requirements for relevant high-risk AI systems shall be part of the risk management procedures established pursuant to that Directive

b) Data and data governance – training, validation and testing data sets shall be subject to appropriate data governance and management practices

c) Technical documentation – the documentation should demonstrate that the high-risk AI systems complies with all the requirements in the regulation and provide authorities with all the necessary information to assess compliance with those requirements

d) Record-keeping – high-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events ("logs") while in operation

e) Transparency and provision of information to users – high-risk AI systems shall be accompanied by instructions that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users

f) Human oversight – human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when using a high-risk AI system

g) Accuracy, robustness and cybersecurity – high-risk AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems; the robustness of high-risk AI systems may be achieved through technical redundancy solutions; the technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and the risks.

References

Ardic, O, J Ibrahim and N Mylenko (2011): "Consumer protection laws and regulations in deposit and loan services – a cross-country analysis with a new data set", January.

Bank of England and Financial Conduct Authority (2019): "Machine Learning in UK financial services", October.

——— (2020): "Minutes: artificial intelligence public-private forum – first meeting", October.

Basel Committee on Banking Supervision (2005): "Compliance and the compliance function in banks", April.

——— (2013): "Principles for effective risk data aggregation and risk reporting", January.

——— (2015): "Corporate governance principles for banks", July.

——— (2018): "Stress testing principles", October.

——— (2019): "Consolidated Basel framework", December.

——— (2021a): "Principles for operational resilience", March.

——— (2021b): "Revisions to the principles for the sound management of operational risk", March.

Consumers International (2013): "In search of good practices in financial consumer protection", February.

Crisanto, J C, C Donaldson, D Garcia-Ocampo and J Prenio (2018): "Regulating and supervising the clouds: emerging prudential approaches for insurance companies", *FSI Insights on policy implementation*, no 13, December.

CRO Forum (2019): "Machine decisions: governance of AI and big data analytics", May.

Economist Intelligence Unit (2020): "The road ahead: artificial intelligence and the future of financial services".

European Banking Authority (2020): "Report on big data and advanced analytics", January.

European Commission (2021): "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts", April.

European Insurance and Occupational Pensions Authority (2019): "Big data analytics in motor and health insurance: a thematic review", May.

——— (2021): "Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the European insurance sector", June.

Financial Stability Board (2017): "Artificial intelligence and machine learning in financial services – market developments and financial stability implications", November.

——— (2018): "Strengthening governance frameworks to mitigate misconduct risk: a toolkit for firms and supervisors", April.

——— (2020): "Regulatory and supervisory issues relating to outsourcing and third-party relationships", November.

Fjeld J, N Achten, H Hilligoss, A C Nagy and M Srikumar (2020): "Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI".

French Prudential Supervision and Resolution Authority (ACPR) (2020a): "Governance of artificial intelligence in finance", June.

——— (2020b): "Governance of artificial intelligence in finance – summary of consultation responses", December.

G20 (2019): "G20 ministerial statement on trade and digital economy", June.

G20 and Organisation for Economic Cooperation and Development (2011): "High-level principles on financial consumer protection", October.

——— (2019): "Compendium of effective approaches for financial consumer protection in the digital age: FCP principles 1, 2, 3, 4, 6, 7, 8 and 9".

German Federal Financial Supervisory Authority (BaFin) (2018): "Big data meets artificial intelligence", July.

——— (2021): "Big data and artificial intelligence: principles for the use of algorithms in decision-making processes", June.

Hong Kong Monetary Authority (2019a): "Consumer protection in respect of use of big data analytics and artificial intelligence by authorized institutions", November.

——— (2019b): "High-level principles on AI", November.

Independent High-Level Expert Group on Artificial Intelligence (2019): "Ethics guidelines for trustworthy AI", April.

International Association of Insurance Supervisors (2019): "Insurance core principles and common framework for the supervision of internationally active insurance groups – update", November.

Luxembourg Financial Sector Supervisory Commission (CSSF) (2018): "Artificial intelligence – opportunities, risks and recommendations for the financial sector", December.

Monetary Authority of Singapore (2018): "Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in Singapore's financial sector", November.

National Association of Insurance Commissioners (2020): "Principles on artificial intelligence (AI)", August.

Netherlands Bank (DNB) (2019): "General principles for the use of Artificial Intelligence in the financial sector", November.

Organisation for Economic Cooperation and Development (2019): "OECD Principles on artificial intelligence", May.

Prudential Regulation Authority (2021): "Supervisory Statement SS2/21: outsourcing and third party risk management", March.

Rudin C (2019): "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", May.

The Joint Forum (2005): "Outsourcing in financial services", February.

UK's Information Commissioner's Office (2020a): "Draft Guidance on the AI auditing framework", February.

——— (2020b): "Guidance on AI and data protection", July.

US regulatory agencies (2021): "Request for information and comment on financial institutions' use of AI, including machine learning", March.

US Treasury (2018): "A financial system that creates economic opportunities: nonbank financials, fintech, and innovation", July.

Veritas Consortium (2020): "FEAT fairness principles assessment methodology", December.

Verma S and J Rubin (2018): "Fairness definitions explained", May.