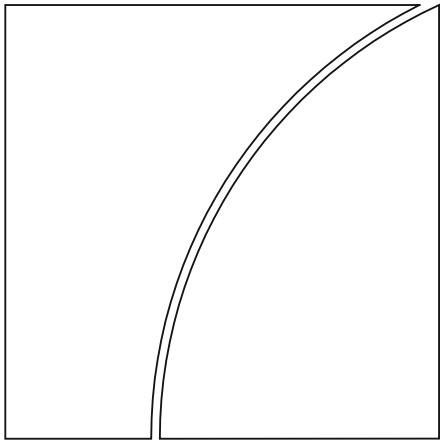
Financial Stability Institute | BIS



FSI Insights

on policy implementation

No 63

Regulating AI in the financial sector: recent developments and main challenges

by Juan Carlos Crisanto, Cris Benson Leuterio, Jermy Prenio and Jeffery Yong

December 2024

JEL classification: C60, G29, G38, O30

Keywords: artificial intelligence, machine learning, corporate governance, risk management, risk modelling

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 Global Media and Public Relations team, please email media@bis.org. You can sign up for email alerts at www.bis.org/emailalerts.htm.

ISSN 2522-249X (online)
ISBN 978-92-9259-822-8 (online)

Contents

Regulating AI in the financial sector: recent developments and main challenges[1]

Executive summary

Financial institutions have been using artificial intelligence (AI) for many years. Three AI use cases are worth highlighting: customer support chatbots; fraud detection, including for purposes of anti-money laundering and combating the financing of terrorism (AML/CFT); and credit and insurance underwriting. Use of AI for chatbots and fraud detection is not new, but the technology has significantly improved in recent years. In terms of credit and insurance underwriting, financial institutions are increasingly using AI for, among others, credit scoring, valuation of collateral and assessing unstructured information from multiple sources to more accurately predict insurance risks and set premiums.

The exponential growth in and accessibility of AI technology is accelerating its use by financial institutions but they seem cautious about generative AI (gen AI). Financial institutions are investing heavily in adopting and implementing AI within their organisations. Much of the increased spending can be attributed to expected wider adoption of gen AI. Financial institutions are experimenting with gen AI to boost operational efficiency and employee productivity. In comparison, gen AI use cases in customer-facing services and high-risk activities are relatively limited. This seems to reflect a cautious approach to gen AI for various reasons, including concerns about customer acceptance and impact; overreliance on third-party model providers; and regulatory uncertainty.

The wider use of AI has the potential to bring transformative benefits to the financial sector but may also exacerbate existing risks. The risks AI poses when used by financial institutions are largely the same risks financial authorities are typically concerned about. These include microprudential risks, such as credit risk, insurance risk, model risk, operational risks, reputational risks; conduct or consumer protection risks; and macroprudential or financial stability risks. Admittedly, AI use may heighten some of the existing risks, such as model risk (eg lack of explainability makes it challenging to assess appropriateness of AI models) and data-related risks (eg privacy, security, bias).

To address AI-related risks, international and national authorities have introduced (cross-) sectoral AI-specific guidance. This guidance outlines policy expectations around common themes. These include reliability/soundness, accountability, transparency, fairness and ethics. More recent guidance has placed increased emphasis on data privacy/protection, safety and security. With the increasing attention on gen AI, sustainability and intellectual property are also being covered in the latest AI guidance. These themes are interconnected and there may be trade-offs between them when developing or upgrading AI guidance. Regardless, the guidance generally allows for a proportionate or risk-based approach to the application of the policy expectations.

The common themes contained in cross-sectoral AI-specific guidance are the same themes emphasised in financial regulations. The common themes in policy expectations are broadly contained in financial regulations covering governance, risk management and consumer protection. This may be the reason why most financial authorities have not issued separate regulations on AI use by financial institutions. Some authorities have issued only high-level principles that reiterate the common themes in

[1] Juan Carlos Crisanto (juan-carlos.crisanto@bis.org), Jermy Prenio (Jermy.Prenio@bis.org) and Jeffery Yong (jeffery.yong@bis.org), Bank for International Settlements; Cris Benson Leuterio (leuteriocs@bsp.gov.ph), Bangko Sentral ng Pilipinas. We are extremely grateful to Iñaki Aldasoro, Gengli Cheng, Leonardo Gambacorta, Ulf Lewrick, Aristides Andrade Cavalcante Neto, Sibel Oezcan, Alain Otaegui, Joe Perry, Brendan Rowan, Vatsala Shreeti, Monika Spudic, Yuta Takanashi and Hanne van Voorden. Graham Austin and Lara Sousa Faria provided helpful research support, while Anna Henzmann provided valuable administrative support.

the cross-sectoral guidance. Other authorities and a few global standard-setting bodies have issued clarifications as to how existing financial regulations apply to AI. So far, among those covered in this paper, only a few authorities have issued regulations specifically addressing AI use by financial institutions.

Nevertheless, AI use by financial institutions may present some unique challenges and hence regulatory or supervisory guidance may be needed in specific areas. Guidance on specific areas can be more important for AI use in financial institutions' core businesses or use cases that present higher risks or significant potential impact on customers. Financial authorities may need to examine existing regulations and, if needed, issue clarifications, revisions or even new regulations in these areas:

- **Governance framework.** The board and senior management of financial institutions are ultimately accountable for their activities, including AI use cases. That said, the use of AI by financial institutions, particularly in their core business activities, would require clear allocation of roles and responsibilities across the entire AI life cycle. Importantly, the governance framework might need to specify the role of human intervention to minimise harmful outcomes from AI systems.
- **AI expertise and skills.** A wider adoption of AI without the corresponding expertise and skills could result in insufficient understanding and ineffective management of the risks to financial institutions and the financial system. Financial authorities may therefore consider clarifying their expectations regarding the expertise and skills envisaged to be in place for financial institutions that plan on expanding AI use in their core business activities.
- **Model risk management.** Heightened model risk can be caused by lack of explainability of AI models. When model risk management guidance is in place, authorities might find it helpful to communicate their explainability-related expectations and provide guidance on the key qualities to consider when selecting explainability techniques and assessing their effectiveness.
- **Data governance and management.** Use of AI by financial institutions can lead to various data-related issues. While many of the relevant elements of data governance/management are captured in existing regulations (eg those for model risk, consumer privacy and information security), financial authorities may want to assess whether these are enough or need strengthening, or whether there is a need to issue guidance that addresses any AI data governance and management-related issues.
- **New/non-traditional players and new business models/arrangements.** To avoid potential regulatory gaps, regulations relevant to new/non-traditional players providing financial services would need to be assessed to determine whether they require adjustments to take account of the cross-sectoral expectations on the use of AI. A similar regulatory assessment might be needed with respect to multi-layer arrangements in providing financial services (eg Banking-as-a-Service) involving AI that may make it challenging for financial authorities to attribute accountability to various players in the ecosystem.
- **Regulatory perimeter – third parties.** The concentration of cloud and AI service providers to a few large global technology firms strengthens the argument for putting in place direct oversight frameworks for these service providers depending on available legal authority. Some jurisdictions have moved in this direction, but the prevalent approach is still relying on financial institutions to manage risks from these third-party relationships.

The presence of various AI definitions across jurisdictions needs to be addressed by international collaboration. The lack of a globally accepted definition of AI prevents a better understanding of AI use cases in the global financial sector and the identification of specific areas where risks may be heightened. As such, international public-private collaborative efforts can be geared towards agreeing on a lexicon for AI and continue working towards regulatory and supervisory frameworks that can adapt to the rapid advancements in AI technology.

Section 1 – Introduction

1. **The artificial intelligence (AI) summer has dawned, prompted largely by the unleashing of Generative AI (gen AI) applications in 2022.** AI can be traced back to the late 1950s, but significant growth in computing power and availability of data accelerated developments only relatively recently. The field of machine learning advanced significantly in the 1990s, while deep learning took off in the 2010s.[2] While AI has caught the general public's imagination for decades, it was only when ChatGPT – a gen AI application – was launched in late 2022 that AI became more readily and publicly accessible. This reignited the interest in AI from the public, businesses – including financial institutions – and national and global authorities.

2. **There is currently no globally accepted definition of "AI" for financial regulatory purposes but there is alignment towards the OECD definition.** This states that *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment"*.[3] IAIS (2024a) considers the OECD definition of AI systems as a useful reference. The definition under the European Union (EU) AI Act converges with the OECD definition but falls short of fully adopting it.[4] Outside of the EU, jurisdictions also have their own slightly different AI definitions but they are generally non-legal, non-prescriptive and non-mandated.[5] This lack of consensus makes it challenging – particularly for firms operating globally – to distinguish what is and what is not AI, as well as the different types of AI. Even at the national level, the intentionally broad definitions of AI may fail to provide a clear differentiation between AI and non-AI systems or may inadvertently capture "basic" statistical models that have been used in the financial industry for many years.

3. **Use of AI by financial institutions preceded the explosion of gen AI applications.[6,7]** Since AI applications have been around for a while, they have been used for various purposes as well. For example, banks may take advantage of opportunities to increase their operational efficiency and facilitate improvements in their risk management by using AI.[8] Insurers have been using AI to facilitate processes such as underwriting, risk assessment and claims management.[9] The exponential growth in and accessibility of AI technology is accelerating the use of AI by financial institutions. Naturally, financial authorities are closely monitoring any potential prudential, conduct and financial stability implications of a wider use of AI in the financial sector.

4. **National authorities in many jurisdictions have introduced cross-sectoral AI-specific policies, but financial authorities have been less active in developing specific regulations.** There were not that many jurisdictions that had cross-sectoral AI-specific policies (ie regulations, guidelines and/or

2 See BIS (2024).

3 See OECD (2024a). A core component of AI systems are AI models, which are used to make inferences from inputs to produce outputs (see Grobelnik et al (2024)). This paper uses the terms "AI system" and "AI model", where appropriate.

4 See, for example, Gulley and Hilliard (2024) for a comparison of different AI definitions.

5 See OECD (2024b).

6 This paper focuses mainly on banks and insurers.

7 Use of AI-enabled tools by financial authorities to support supervisory work – so-called suptech tools – also precedes gen AI developments. While financial authorities face the same risks in the use of AI as financial institutions, this paper focuses only on the latter. See also Prenio (2024) and Aldasoro, I, L Gambacorta, A Korinek, V Shreeti and M Stein (2024) (2024).

8 See BCBS (2022).

9 See Ladva and Grasso (2024).

frameworks that apply to the use of AI across multiple industries or sectors) a few years back.[10] However, a large number of jurisdictions now have different types of policies that cover AI either specifically or in the context of data protection, cyber security and consumer protection requirements, among others.[11] Many of these policies have been issued by national authorities, either in the form of binding legislation or non-binding guidance. Only in a small number of jurisdictions have financial authorities issued AI-specific regulations. Interestingly, the majority of respondents to an OECD survey do not plan to introduce new regulations on AI use in finance in the near future.[12] This could be explained by the fact that risks arising from AI are not new[13] and are already addressed in existing financial regulations, and financial authorities are also generally taking a technology-neutral approach when issuing regulations.

5. **While financial authorities generally follow a technology-neutral approach[14], they may need to enforce relevant provisions of cross-sectoral AI-specific policies.** Broadly speaking, under a risk-based approach, supervisors require assurance that financial institutions understand the risks that they are taking and have proper governance, risk management and controls to identify, monitor, manage and mitigate these risks. However, financial authorities may need to clarify how existing regulations apply when implementing relevant provisions of cross-sectoral AI-specific policies. Moreover, there may be a need to strengthen existing regulations or issue new regulations on specific areas to respond to the unique and practical enforcement challenges given the characteristics of AI and how they are deployed.

6. **This paper identifies the practical challenges involved in enforcing regulatory expectations on AI and specific guidance that may be helpful in addressing some of those challenges.** Many of the existing papers on regulation of AI typically describe the regulatory requirements and expectations but fall short of describing how these could be implemented in practice. Our paper aims to fill this gap by updating Prenio and Yong (2021) and looking at newer guidance, particularly that issued in Brazil, China, the EU, Qatar, Singapore, the United Kingdom and the United States. Section 2 starts by providing an overview of AI use cases in the banking and insurance sector. It is based on desktop research and discussions with financial institutions. Section 3 outlines the common themes of cross-sectoral AI-specific guidance and the emerging policy framework for the use of AI in finance. Section 4 discusses the practical issues in enforcing some of the themes or expectations. Here, the paper tries to anchor the discussion on concrete use cases, ie credit and insurance underwriting. These are the financial-sector specific use cases that have been identified as "high-risk" under the EU AI Act. Section 5 concludes.

Section 2 – Overview of AI use cases in the financial sector

7. **Financial institutions are investing heavily in adopting and implementing AI within their organisations.** The large spending suggests that financial institutions are expecting to benefit significantly from their AI investments.[15] Such transformative changes could profoundly alter how financial institutions conduct their business activities, and this alone should warrant closer supervisory scrutiny. Statista estimates that spending by the financial sector on AI will increase from USD 35 billion in 2023 to USD 97

[10] See Prenio and Yong (2021).

[11] See OECD (2024b); Stanford University (2024) analyses legislation in 128 countries during the period 2016–23 and finds that, in total, these countries have passed 148 AI-related bills and 32 have enacted at least one AI-related bill.

[12] ibid. The OECD survey took place in the first quarter of 2024 and involved 49 OECD and non-OECD jurisdictions.

[13] See IAIS (2023b).

[14] A technology-neutral regulatory/supervisory approach does not differentiate between the different technologies, whether AI or not, that a firm may use.

[15] According to the World Intellectual Property Organization (WIPO) ranking in 2024, Ping An (one of the largest insurers in China) ranked second globally with 1,564 generative AI patent applications. Ping An is reported to have a technology team of more than 20,000 technology developers and over 3,000 scientists as of 30 June 2024.

billion in 2027.[16] Much of the increased spending can be attributed to expected wider adoption of gen AI. The banking sector's spending on gen AI alone is expected to increase from USD 3.86 billion in 2023 to almost USD 85 billion in 2030. Much of this AI-related spending is on headcount and IT infrastructure. A study[17] found that major banks are increasing AI talent headcount even though they are cutting headcount elsewhere, suggesting expected AI productivity gains that can replace human resources. McKinsey (2024) estimates that gen AI could add between USD 200 billion and USD 340 billion in value annually, or 2.7% to 4.7% of total industry revenues, mainly through increased productivity.[18]

AI use cases: banks and insurers

8. **There are different ways to categorise financial institutions' AI use cases.** For example, use cases can be categorised in terms of the business value chain[19], job functions[20], risk types/levels[21] or types of AI algorithms[22]. As AI use cases by banks and insurers are expanding very quickly, it is difficult to summarise or identify the most prevalent use cases. This paper provides a point-in-time snapshot of how financial institutions are using AI based on feedback from selected industry players and through industry surveys.[23]

9. **This paper classifies AI use cases based on their purpose/objective while recognising that it is difficult to generalise AI use across all financial institutions.** Use cases may vary from one financial institution to another due to heterogeneity in terms of different sizes and types of firms (eg digital banks/insurers).[24] Some firms are taking a more cautious approach, using AI predominantly for back office, operational purposes, while others are more open to the use of AI in core business and revenue-generating activities. Nevertheless, reported in-production use cases for core, external-facing business activities are not prevalent yet. From a regulatory perspective, it should be acknowledged that AI has the potential to be used across all business activities and, importantly, has the potential to become the "norm" in supporting all financial services activities. Table 1 provides examples of actual AI use cases by selected banks and insurers:[25]

16 See Statista (2024).

17 See Evident (2024).

18 JPMorgan Chase (2024) estimates the value of its AI deployment to be around USD 1 to 1.5 billion in terms of productivity improvements and cost reduction, citing an example of know-your-client file processing. They expect to increase the number of files processed from around 155,000 in 2022 to 230,000 in 2025 but with 20% less staff needed to do so. DBS Singapore has deployed over 800 AI models across 350 use cases and estimated an economic impact exceeding SGD 1 billion in 2025.

19 See BCBS (2024), The Economist Intelligence Unit (2022).

20 See Accenture (2024).

21 See European Parliament (2024) and MIT's AI risk repository.

22 See EIOPA (2024).

23 See IIF-EY (2023), NVIDIA (2024).

24 See BIS (2024).

25 It is acknowledged that some of the use cases may be classified differently under different objectives. This table is not intended to provide distinct demarcation of the various use cases; rather, it is intended to illustrate the range of use cases that support key business objectives.

Banks' and insurers' use of AI — Table 1

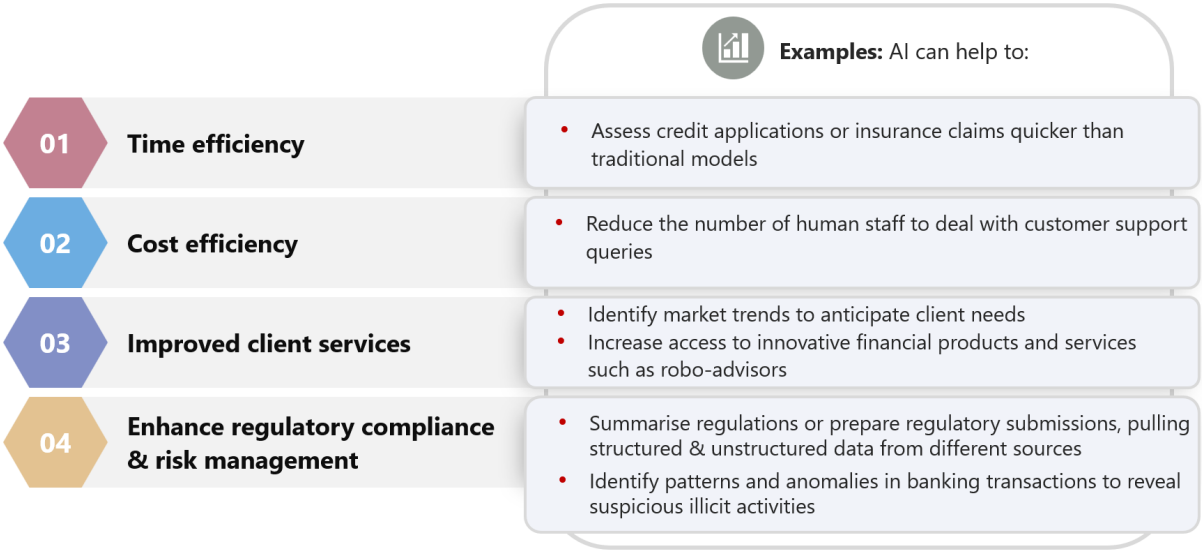
Objective	Use case	Description	Example
Improve productivity and efficiency	Internal administrative tasks	• Summarise documents or internal meetings • Classify documents	Standard Chartered Axa Secure GPT
	Customer support[1]	• Chatbots to respond to customer queries • Automate email response to clients	Bank of America Erica customer chatbot DBS CSO Assistant Ping An[2] AI service representatives JPMorgan Chase email classification system
	Human resource management	• Virtual reality training on customer interaction	Bank of America
	Coding	• Facilitate coding of IT applications	Goldman Sachs
	Insurance claims	• Use of AI to estimate property damage	MS&AD use of Tractable
	Reinsurance claims	• Automate identification of reinsurance claims	Zurich Catastrophe Insurance Agent
Support regulatory compliance and risk management	Regtech[3]	• Analyse regulatory requirements including through regtech	Citi use of gen AI to read US banking capital rules
	AML/CFT	• Detect suspicious activities	HSBC AML AI detection tool
	Fraud detection[4]	• Real-time monitoring of unauthorised credit card transactions	Société General MOSAIC fraud detection AI tool
	Cyber security[5]	• Enhance cyber resilience	Barclays
Enhance core business/revenue-generating activities	Credit underwriting	• Data analysis to determine loan eligibility	MUFG DBS
	Insurance underwriting	• Accelerate processing of insurance applications	ICICI Prudential

[1] Customer support may also be considered as a revenue generating tool as retained and satisfied customers can purchase more services or stay loyal to a firm. [2] Ping An reportedly handled around 870 million interactions (80% of its customer service queries) using its AI service representatives in the first half of 2024. [3] Hong Kong Monetary Authority (2022) provides an overview of AI-based Regtech solutions, implementation challenges and sample use cases. [4] Reinsurance Group of America (2024) found that 48% of their surveyed insurers suffered AI-related fraud such as falsified medical or death records, deepfake or voice cloning. [5] Bank of England (2024) reported that 37% of surveyed UK financial services firm use AI for cyber security purposes. BIS (2024) outlined its Innovation Hub's projects in AI, ranging from AML/CFT to extracting climate-related data.

Source: FSI authors.

10. **Financial institutions can use AI to do things quicker, cheaper and better and, importantly, to do things that humans cannot do with the accuracy and speed that AI can deliver.** Supervised and unsupervised AI models can be used to make predictions by learning from patterns or trained to look for patterns themselves. Such capabilities can offer tremendous opportunities to financial institutions and may significantly transform financial services. Examples of use cases for each desired outcome are provided in Figure 1:

Main desired outcomes of AI use cases in financial institutions Figure 1



Source: FSI authors.

11. **From a regulatory compliance standpoint, AI has the potential to support prudential objectives.** Regulatory technology, or regtech, refers to applications that financial institutions can use to meet regulatory requirements. These include technology solutions that help financial institutions comply with regulatory reporting, anti-money laundering and combating the financing of terrorism (AML/CFT) and calculation of regulatory capital, among others. Rapid advancements in AI offer new capabilities for financial institutions to fulfil regulatory requirements in a more effective and efficient way. This may improve the safety and soundness of the financial sector as banks and insurers become better able to comply with regulatory requirements.[26]

12. **Most financial institutions' AI use cases reviewed for this paper are for internal operational efficiency purposes, and less for core business activities.**[27] According to BCBS (2024), some banks have been cautious in adopting AI due to uncertainties surrounding regulatory expectations related to accountability, ethics, data privacy, fairness, transparency and explainability, particularly for consumer-related applications. Gen AI use cases in customer-facing services and high-risk activities are limited, while some banks are experimenting with gen AI to boost operational efficiency and employee productivity. OECD (2023) attributes the slow implementation of gen AI in financial markets to strict regulations and potential adverse impact on customers. Concerns over data sovereignty and globally dispersed data (NVIDIA (2024)), as well as legacy IT infrastructure (KPMG (2023)), also pose significant challenges to rapid deployment of gen AI. An industry study, IIF-EY (2023), reported that firms expect gradual deployment of gen AI to limit any potential negative impact on external stakeholders while the technology matures further.

13. **The use of AI for customer support is common.** Chatbots are not new features in financial services, but the technology has improved significantly over the years. The main motivations for the use of AI-powered chatbots are to cut cost by reducing human interaction time and improve customer

[26] FATF (2021) describes how AI can be used for AML/CFT purposes. Oracle (2024) cites a McKinsey study reporting that AI can improve identification of suspicious activities by 40%.

[27] HKMA (2024b) reports that most of the surveyed financial institutions in Hong Kong SAR have adopted or are planning to adopt AI for operational automation and document processing.

experience by providing 24/7 support.[28] At a basic level, chatbots can provide information about a financial product. More sophisticated chatbots are used to offer personalised financial services such as tracking of personal spending. Some advanced chatbots can even execute financial transactions such as loan applications.[29] Chatbots might be an area of focus for financial regulators because of their wide and growing reach. CFPB (2023) estimates that 37% of the US population interacted with a bank's chatbot in 2022.[30] As human-interfacing AI technology improves further, for example by allowing people to converse verbally with a chatbot in different languages, the use of chatbots by financial services firms can be expected to increase.

14. **Another AI use case in the financial sector is to detect money laundering/terrorism financing and fraud activities.** Similar to chatbots, the use of AI for these purposes is not new. What is new is the more widespread use of AI tools by financial institutions, and their improving accuracy.[31] Such AI tools assist in flagging the rapid movement of money into different accounts, or transactions that significantly deviate from anticipated patterns. The tools are becoming more effective in identifying suspicious individuals, mule accounts and organised groups that exploit the vulnerabilities in rules-based systems. The tools are reducing the number of alerts or false positive cases, freeing up time to allow institutions to carry out comprehensive investigations on legitimate cases. Another notable and related example intersecting with AML/CFT is payments fraud emanating from digital financial services. An AI fraud management system can be used to prevent or detect suspicious payments, and promptly alert financial institutions of unusual transactions. This enables financial institutions to review and decide whether to approve or reject the seemingly irregular payments. The solution can also adjust to unique customer behaviours and evolve along with the business.

15. **Underwriting is an area where AI is increasingly being used, with some insurers appearing to be more advanced than banks.** Insurance underwriting can rely on simple questionnaires to assess the insured risks (for example, for life insurance products), or it may involve complex risk assessments that require physical examination of the insured property and written assessments from underwriters (for example, commercial property insurance). AI, and in particular gen AI, can be useful to assess unstructured information from multiple sources in insurance underwriting processes to more accurately predict risks and set premiums. In banking, machine learning has been used for many years in credit underwriting.[32] It is used for credit scoring, valuation of collateral, calculating the interest rate to charge and personalisation of loan offers, sometimes with the aid of synthetic data (ie data artificially generated by using, for instance, algorithms).[33]

16. **Use of AI for underwriting can help to address and mitigate some of the challenges financial institutions face.** For credit underwriting, this includes high operational cost due to time-consuming and manual processes, risk of fraud and subjectivity. AI could greatly enhance credit scoring by making use of unstructured data (ie non-traditional financial information).[34] Insurance underwriting processes vary depending on the complexity of the coverage and extent of risks insured. Commercial risks (eg marine insurance) require assessment of voluminous reports from different sources (eg vessel

[28] Forrester (2023) estimates that a chatbot in a stylised financial institution reduced human interaction handle time by up to 30%.

[29] DBS digibot can execute loan applications with instant funds transfers to successful applicants.

[30] By June 2023, Bank of America's chatbot, Erica, had recorded 1.5 billion interactions with more than 37 million clients since its launch in June 2018. It is reported that Bradesco's chatbot answers 283,000 questions each month with a 95% accuracy rate.

[31] HSBC estimates that its AI AML tool identifies two to four times more suspicious activities than its previous system, while reducing the number of alerts by 60%, thus allowing more time for its human investigators to review genuine suspicious cases. The tool also allows identification of criminal networks.

[32] See BIS (2024).

[33] Betterdata.ai explains how synthetic data can be used to create hypothetical data sets covering different credit behaviours and profiles that can be used to train AI systems without biases that may be present in actual data sets.

[34] See BIS (2024).

information, inventory, shipping contract). Use of AI can automate underwriting, provide the ability to analyse large volumes and varied forms of data and improve identity verification, while at the same time enhancing customer experience. In insurance, AI, and especially gen AI, can offer capabilities previously not available in terms of ability to process large volumes of text data.[35] In general, AI can reduce underwriting cost, support financial inclusion[36], and enhance efficiency (eg faster approval turnaround time).

Risks arising from banks' and insurers' AI use cases

17. **While the adoption of AI by banks and insurers offers significant benefits, it also exposes these institutions to a range of risks that require careful management.** There have been many reports on the risks arising from the use of AI. Tables 2 to 5 provide a non-exhaustive list of such risks:[37,38]

Microprudential risks — Table 2

Risk type	Description/example
Credit risk	• Underestimation of probability of default or risk of loss due to inaccurate data inputs
Model risk	• Inaccurate model output due to the model not capturing changes to the nature of the data input[1] • Lack of model explainability hinders the ability to assess its conceptual/technical soundness • Inaccurate model output due to overfitting or underfitting; that is, the model output cannot generalise to other conditions or circumstances, or it is too simplistic and hence fails to capture the underlying patterns in the data • Hallucination, inconsistent responses and dependency on data quality[2] • Overestimation of the capabilities of AI models, leading to misuse of such models beyond their capabilities • AI models may not produce reliable predictions if they are not trained with the most recent information available
Insurance risk	• Underpricing of insurance policies due to AI models trained on historical data not capturing latest developments (eg new disease outbreaks)
Cyber risk	• Firms may be more vulnerable to cyber attacks due to increased contact points with multiple external service providers and increased IT interconnectivity with multiple systems • Inadequate access control may result in unauthorised access to training data and AI model • AI models may be susceptible to data poisoning attacks that alter the training data sets for malicious purposes • Threat actors could "steal" an AI model by constructing a functionally equivalent model through querying a model iteratively
Other operational risk	• Firms with legacy IT systems may add complexity to their IT architecture, thus increasing potential operational risks arising from IT failures

[35] The measurable impact includes reduction of the approval process time by more than two days and a 94% accuracy rate in credit analysis calculations. See Marsch & McLennan Companies (2019).

[36] See Aldasoro, Gambacorta, Korinek, Shreeti and Stein (2024).

[37] See BCBS (2024), Bank of England (2022), ECB (2024), FSB (2017, 2024), IAIS (2023b, 2024a, 2024b (forthcoming)), IMF (2023), UK Government (2024), US Department of the Treasury (2024).

[38] The OECD collects data on AI incidents, which can be accessed here: OECD.

	• Increased use of third-party services (data providers, AI model providers) could lead to dependency, disruption of critical services and lack of control of processes, which may be exacerbated by vendor lock-in risk and increased market concentration • Quick obsolescence of risk controls due to rapid updates by AI systems
Reputational risk	• Operational failures, potentially due to overdependency on third-party providers, can damage public trust and confidence • Adverse publicity due to unfair treatment of customers or regulatory penalties can erode reputation of firms
Strategic risk	• Financial institutions partnering with other firms may lose control over critical functions such as business origination and customer relationships, potentially resulting in significant liquidity issues and financial instability if those partners redirect business or alter key processes
Legal risk	• Firms may be liable for copyright infringement due to unauthorised use of copyrighted data in training AI models • Firms may be exposed to legal liability due to inaccurate or inappropriate response provided by customer-facing AI tools

[1] See What Is Model Drift? | IBM. [2] See FSOC (2023).

Sources: See footnote 37.

Conduct/consumer protection risks

Table 3

Risk type	Description/example
Unfair treatment of customers	• Exploiting characteristics of vulnerability of consumers to charge unfair prices • Arriving at discriminatory decisions based on biased data or personal information in alternative data used to perpetuate bias • Financial exclusion of perceived high-risk customers
Price collusion	• Collusive pricing strategy implemented by automating price adjustments based on pricing changes by competitors

Sources: See footnote 37.

Macroprudential/financial stability risks

Table 4

Risk type	Description/example
Herding behaviour	• Amplification of procyclical behaviour due to the use of similar data sets and AI models by multiple financial institutions • AI outputs may contribute to market participants' conclusions being systemically biased, leading to distorted asset prices or increased price correlations
Interconnectedness and concentration	• Increased interconnectivity amongst firms from highly concentrated AI third-party providers could result in systemic risk if those third parties suffer from cyber attacks or operational failures, affecting multiple financial institutions and markets simultaneously
Opacity and complexity	• Limits to the explainability of certain complex AI models can result in risk management challenges, as well as lesser financial institution and supervisory insight into the build-up of systemic risks

Sources: See footnote 37.

Other risks

Table 5

Risk type	Description/example
Market competition risk	• The high cost of developing and maintaining AI technologies may limit their adoption to larger financial institutions, potentially increasing the market power and systemic importance of these firms, while making it difficult for smaller firms to compete
Data privacy risk	• AI models may be manipulated to leak personal or sensitive information used in training and using the models
Environmental risk	• Heightened use of AI will increase energy demand, which may contribute to climate change[1]

[1] UK Government (2024) estimates that in 2026, computing power for AI will consume roughly the same amount of electricity as smaller European countries such as Austria or Finland.

Sources: See footnote 37.

18. **AI can be a double-edged sword for cyber resilience.**[39] AI can significantly strengthen cyber security by proactively detecting threats (including AI deepfakes) and identifying vulnerabilities. Through the analysis of large volumes of (historical) data, AI can help to identify trends as well as unusual patterns that may indicate cyber threats or forecast potential cyber attacks.[40] Gen AI has the potential to take these techniques to a new level through more advanced capabilities such as realistic simulation attacks and real-time adaptive cyber security posture. At the same time, cyber criminals can use similar AI tools to conduct more sophisticated cyber attacks through, for instance, targeting vulnerabilities in underlying models or data or generating realistic fake profiles to be used in social engineering attacks. These can be much harder to detect since they can also be adapted in real time and automated at great scale.[41] In a 2024 global cyber security survey, the majority of respondents believed that in the next two years gen AI would provide overall cyber advantage to attackers, while a third responded that the situation would be balanced between attackers and defenders.[42,43] Regulators are increasingly focusing attention on the use of AI to exploit cyber vulnerabilities of firms.[44]

19. **Consolidation of AI service providers within big techs is a particular concern for both the industry and regulators, as this trend may expose financial institutions to heightened concentration risks.** Big techs are cementing their foothold as they dominate the AI industry and influence the research on AI (West (2023) and Ahmed et al (2023)). Their access to vast quantities of data, the computational power to process them, and expertise to build the AI systems has collectively given them the first-mover advantage. These developments are attracting closer supervisory scrutiny as they can give rise to microprudential and financial stability risks. In 2023, the FSB published a toolkit for financial institutions and financial authorities to manage and oversee third-party risks.[45] In 2024, the Federal Trade Commission launched an investigation into gen AI investments and partnerships between AI companies and major

39 See Aldasoro, Doerr, Gambacorta, Notra, Oliviero and Whyte (2024).

40 BOE and FCA (2024) found from their industry survey that the highest perceived benefits of AI include its use for cyber security.

41 See US Department of the Treasury (2024).

42 See World Economic Forum (2024).

43 UK Government (2024) concludes that currently, there is not yet any substantial evidence suggesting that general purpose AI can automate sophisticated cyber security tasks.

44 New York Department of Financial Services (2024b) provides guidance to financial institutions on how to manage cyber security and related risks arising from AI.

45 See FSB (2023).

cloud service providers (Box 1 explains the use of gen AI in financial services). BCBS (2024) noted that banks' increasing reliance on third-party technology services introduces cyber risks and potential systemic vulnerabilities. IAIS (2024a) highlighted the importance of insurers regularly assessing their reliance on AI service providers that may pose a risk to their business, noting the potential implications of a concentrated market of AI providers. ECB (2024) highlighted how technological penetration (use of AI applications by a large number of firms) and supplier concentration can give rise to systemic risk.

Box 1

Gen AI in financial services

Gen AI refers to AI applications that can generate new content, including text, images or music, from a natural language prompt.① It relies on machine learning models, mainly deep learning, that mimic the learning and decision-making of the human brain. These models work by identifying and encoding the patterns and relationships in enormous amounts of data, and then using that information to understand users' natural language requests or questions and respond with new content.

Gen AI applications are becoming more accessible to financial institutions. Many existing cloud service providers of financial institutions have expanded their offerings to include gen AI applications. At the same time, big techs continue to dominate the gen AI market, owning the majority of foundation models②, ie models that are trained on broad data sets and can be used for a wide range of tasks including gen AI applications. The very high cost③ of training foundation models can be a barrier to entry for smaller firms.

The technical performance of AI models is rapidly improving, surpassing human capabilities according to a study④, including in gen AI use cases. Nevertheless, the foundation models that underpin many gen AI use cases in the financial sector require adjustments to make them fit for purpose, as these models are trained on large data sets, and are intended for a wide range of use cases. To make gen AI outputs more relevant for financial institutions, a technique that can be used is called "retrieval-augmented generation" (RAG).⑤ Through RAG, firms can control the context of a foundation model using its own information or data.

Despite the increasing attention on gen AI and its potential to further increase the benefits indicated in paragraph 11, there have not been widespread use cases by banks and insurers for revenue generation purposes. Insurers seem to have more gen AI use cases than banks. This is probably because insurance products involve more unstructured data than banking products. Insurance products are essentially financial contracts that are very heterogeneous, containing different terms and conditions (precise definition of insured events, exclusions, etc). Moreover, the underwriting and claims management of insurance products may require large amounts of data from different sources. As such, insurance-related activities lend themselves better to the use of gen AI. For example, gen AI can be used to help human underwriters more quickly identify appropriate policies and terms based on the information provided by the prospective customer.

Firms seem particularly cautious in using gen AI for customer-facing use cases. This can be attributed to the following:

- heightened risk exposures, for example potential mis-selling or provision of wrong advice;
- the high bar needed to fulfil relevant regulatory requirements, for example the need to validate the model results;
- firms' own internal risk management policy, for example customer information disclosure requirements before concluding a transaction;
- lack of clarity on the party ultimately accountable if the model results are wrong;
- lack of consumer trust to interact with gen AI; and
- overreliance on third-party model providers.

The risks posed by gen AI are mainly an extension or amplification of existing model risks. Compared to other AI models, gen AI gives rise to unique risks related to anthropomorphism, treating the AI models as though they have human-like qualities. Overestimating the capabilities of gen AI is becoming more perennial as publicly accessible gen AI applications offer more human-like features such as voice and visual conversation. Users may come under the

false impression that such models can actually think, reason or even display emotions. Perez-Cruz and Shin (2024) explain that gen AI models are susceptible to reasoning errors and cognitive limit. BCBS (2024) highlights the potential of gen AI to hallucinate⑥ by generating responses that are inaccurate or inappropriate, and by producing different responses over time, even when given similar questions or prompts. This is because gen AI outputs are characterised by randomness. Such risks are contributing to the cautious rollout of customer-facing gen AI use cases in financial services.⑦

The "democratisation" of gen AI, making the technology available to virtually everyone, has accelerated financial institutions' beefing-up of their internal AI governance and risk management policies. Some firms have decided to ban the use of gen AI while they figure out how guardrails can be put in place for its safe and responsible use. New governance structures are emerging, for example, formation of senior management committees to screen gen AI use cases under a risk-based approach. Use cases that involve complex models and autonomous decision-making by the model and that are customer-facing/impactful will attract greater scrutiny and risk controls. Firms are starting to establish a use case and risk registry to systematically monitor their gen AI activities as well as "AI factories" with dedicated staff working with all the necessary infrastructure and data layers in one place, including gen AI models, both open source and third-party models accessed via cloud APIs.

① See BIS (2024). ② Stanford University (2024a) reports that 97 out of 163 foundation models released between 2019 and 2023 are owned by four big techs – Google, OpenAI, Meta and Microsoft. ③ Stanford University (2024a) estimates that the training of OpenAI's GPT-4 and Google's Gemini Ultra cost around USD 78 million and USD 191 million respectively. ④ See Stanford University (2024a). ⑤ See What is RAG? - Retrieval-Augmented Generation AI Explained - AWS (amazon.com). ⑥ A study estimates that the hallucination rate of large language models (LLMs) ranges between 1.4% and 4.2%. ⑦ See Calabia (2024) for a thorough discussion on the benefits and challenges of gen AI for financial services and financial regulation.

20. **Anticipated widespread use of AI without adequate supervisory oversight and sound risk management practices in firms could pose threats to the safety and soundness of the financial sector.** Although it is uncertain how AI applications will evolve,[46] it is plausible that the use cases within the financial services industry will continue to expand as the technology becomes more accessible, and it does not take much imagination to see how AI could become ubiquitous in financial institutions' technology infrastructure. Firms may accelerate adoption of AI to improve productivity and make business gains. Even late adopters, or even resisters, might be pushed to adopt AI due to the "fear of missing out" compared to their competitors. As such, financial sector regulators may need to anticipate a future where AI systems become integral across the entire value chain of financial services activity. The risks arising from such widespread deployment need to be properly understood so that regulators can ascertain if their existing toolkit will remain fit for purpose.

Section 3 – Overview of cross-sectoral AI-specific guidance

21. **Multilateral groups and international organisations are giving priority to the development of AI policy.** The G20 has emphasised the need for human-centric and trustworthy AI. These objectives were reflected in the AI Guidelines adopted in 2019[47], which largely built upon the OECD AI Principles.[48] The G7 has also been actively coordinating a policy response to AI developments, including gen AI, and a milestone was achieved in December 2023 with the endorsement by G7 leaders of "the Hiroshima AI Process Comprehensive Policy Framework".[49] This provides guiding principles and a code of

[46] UK Government (2024) highlights disagreement within the global AI scientific community on whether AI technology will continue to develop and advance.

[47] See G20 (2019).

[48] This position was also reflected in subsequent G20 Leaders' Statements in 2019 (Japan), 2020 (Saudi Arabia), 2021 (Italy), 2022 (Indonesia), 2023 (India) and 2024 (Brazil). See the Center for AI and Digital Policy (CAIDP).

[49] The Hiroshima AI Process was launched in May 2023. More details can be found on its official website: soumu.go.jp.

conduct aimed at promoting safe, secure and trustworthy advanced AI systems.[50] More universally, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has issued recommendations on the ethics of AI, which were adopted by all 193 UN member states in November 2021.[51] Moreover, the UN adopted its first ever resolution on AI, emphasising its role for sustainable development, in March 2024[52] and published its final report on global AI governance in September 2024.[53]

22. **The OECD AI Principles are a key reference point when developing AI policy at the national level.** These non-binding principles were initially adopted in 2019 and updated in 2024. The AI Principles guide the development of trustworthy AI systems based on value-based principles such as inclusiveness, sustainability and well-being; human rights and democratic values including fairness and privacy; transparency and explainability; robustness, security and safety; as well as accountability. They also provide policymakers with recommendations for effective AI policies.[54] The 2024 update aims to ensure that the AI Principles continue to be technically accurate and reflect technological developments, particularly the growing importance of gen AI.

23. **Jurisdictional policy approaches to deal with AI can be broadly categorised as principles-based and rules-based approaches.** Jurisdictions opting for the former largely rely on non-binding principles and/or voluntary commitments generally supported by technical standards and/or cross-sectoral regulations (eg Singapore, United Kingdom, United States). While this approach recognises the risks and challenges brought about by AI, these jurisdictions consider it too early to regulate AI in a forceful way given the ongoing evolution of this technology. Jurisdictions opting for a rules-based approach have issued or are in the process of issuing AI legislation (eg Brazil, China, European Union and Qatar). This approach seeks to provide regulatory clarity to facilitate the safe advancement of this technology and the legal powers for enforcement against unlawful AI deployment. Some of these jurisdictions consider it imperative to protect consumers' rights from potential harms.

24. **AI guidance generally allows for proportionate or risk-based application.** The concept of proportionality in the context of AI policy is informed by the need to avoid imposing unnecessary or disproportionate costs and/or burdens on businesses and regulators. The policy measures vary in stringency based on the outcomes that an AI system is likely to generate rather than having uniform rules applied to the technology itself or its applications.[55] The rationale of a risk-based approach to AI is to foster innovation without compromising the development of trustworthy AI systems. By focusing on the potential risks associated with different AI applications, this approach aims to ensure that policy efforts aiming at minimising harms and promoting responsible AI systems are efficient and effective. The approach can address concerns surrounding inadvertent wide scope of what is considered an "AI system" by excluding non-consequential AI use cases (eg summarisation of internal meeting minutes) from regulations.

25. **Regardless of the policy approach taken, cross-sectoral AI-specific guidance continues to cover common themes and highlight additional ones.** Prenio and Yong (2021) identified five common

[50] See G7 (2023a,b,c).

[51] See UNESCO (2022).

[52] See UN (2024a).

[53] This was put together by the UN Secretary General's High-level Advisory Body on AI. See UN (2024b) and www.un.org/en/ai-advisory-body/about.

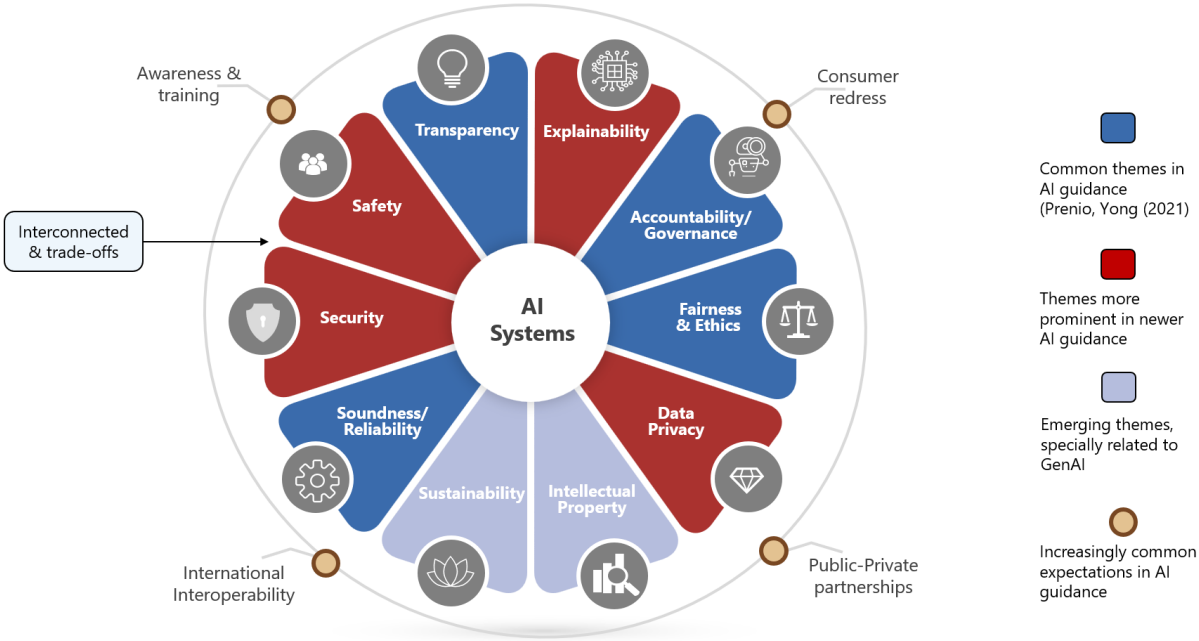
[54] Recommendations for policymakers include investing in AI research and development; fostering an inclusive AI-enabling ecosystem; shaping and enabling an interoperable governance and policy environment for AI; building human capacity preparing for labour market transition; and international cooperation on trustworthy AI.

[55] Regulatory treatment may vary by the type of client (wholesale versus retail) of financial institutions. In the insurance sector, the use of AI by reinsurers with respect to their clients, primary insurers, may not attract the same level of regulatory scrutiny as AI use cases that impact retail policyholders.

themes: reliability/soundness, accountability, transparency, fairness and ethics.[56] More recent AI-related guidance continues to cover these themes except for ethics, explicit coverage of which is somehow less evident. Newer AI guidance consistently highlights additional themes such as security, safety, explainability and data privacy. It also provides some more concrete guidance as to how authorities expect these themes to be addressed. With the increasing attention on gen AI, sustainability and intellectual property are also being covered in more recent AI guidance. Additionally, newer guidance is consistently featuring topics such as consumer redress; awareness and training; international interoperability; and public-private partnerships. The following paragraphs review the above-mentioned common themes, as well as additional topics and features that have been highlighted in recent AI guidance.

26. **The common themes are interconnected and there may be trade-offs between them when developing or upgrading an AI policy framework.** Transparency, for example, is considered as enabling the assessment of the other themes; that is, without transparency, it would be challenging to assess the reliability of an AI model and to enforce accountability, fairness and ethics. At the same time, there might be a trade-off between reliability and transparency (including the concept of explainability), as the precision of an AI system may require more data inputs or parameters, such as in the case of gen AI, thus making the model more complex. This, in turn, may affect decisions around fairness. To operationalise policy expectations across common themes, Aldasoro et al (2024) provide a framework for regulating gen AI and AI agents in finance,[57] building upon core activities in dealing with AI (ie govern, map, measure and manage) and the main stages in the AI value chain (ie design and training; deployment and usage; and longer-term diffusion). Figure 2 summarises the common themes in cross-sectoral AI-specific guidance.

Themes in cross-sectoral AI-specific guidance — Figure 2



Source: FSI authors.

[56] Different authorities may use other terms to characterise similar concepts or may group certain concepts together (eg reliability/soundness under fairness). Prenio and Yong (2021) use authors' judgment in naming or distinguishing the different concepts.

[57] Aldasoro, I, L Gambacorta, A Korinek, V Shreeti and M Stein (2024) define AI agents as AI systems that build on advanced LLMs such as GPT-4 or Claude 3 and are endowed with planning capabilities, long-term memory and, typically, access to external tools such as the ability to execute computer code, use the internet, or perform market trades.

Transparency and explainability

27. **One aspect of transparency relates to internal transparency, which refers to explainability, interpretability and auditability of AI models.** An explainable AI model makes transparent how it arrived at a certain outcome. Explainability is especially emphasised, even more than reliability, when the model's use may have a significant potential impact on customers or the public. It is therefore concerning that, for gen AI models, Stanford (2024b)[58] found that most foundation models remain opaque. An auditable AI model requires proper documentation of its design, processes and the data used. Another aspect of transparency is interpretability. NIST (2023) distinguishes explainability and interpretability as follows: the former answers the question of "how" and the latter answers the question of "why" a decision was made by the AI system. In essence, explainability, interpretability and auditability involve internal disclosure or transparency particularly to the board and senior management so they can better understand the risks and implications of AI use.

28. **External transparency of AI systems towards customers is also important.** This is driven by the fairness objective and includes disclosing to customers when they are interacting with AI (eg their data are being used by AI); the use of AI-driven decisions that affect them; and consequences of AI-driven decisions on them. More recent guidance mentions providing an explanation about the decision, including the "logic" or "rationale" of the decisions and the contribution of the AI models to these decisions. The guidance often states that such disclosures should be in the form of plain and easy-to-understand information. Box 2 provides an overview of emerging supervisory expectations on explainability of AI systems.

Box 2

Emerging high-level expectations on explainability in AI systems

AI guidance and model risk management (MRM) frameworks are currently the primary tool to manage and mitigate AI-related risks, including opacity and lack of explainability. Building upon the experience of implementing MRM frameworks in the financial sector, high-level expectations are emerging to foster explainable AI systems. One of the most widely recognised efforts in this area is the NIST's four non-binding principles of explainable AI (NIST (2021)). According to these principles, an explainable AI system should:

(i) provide supporting evidence or reasons for its outputs and processes (supported decision-making);

(ii) offer explanations that are understandable to the intended users (understandable explanation);

(iii) accurately reflect the reasoning behind the output and faithfully represent the system's processes (explanation accuracy); and

(iv) only operate under conditions for which it was designed and when it reaches sufficient confidence in its output (capability limits).

The following paragraphs assess the extent to which these principles are explicitly or implicitly incorporated into the AI guidance and MRM frameworks under review.

The expectation to provide supported decision-making explanations is included in AI guidance but not always explicitly stated in MRM frameworks. That said, this can be inferred in specific contexts such as model validation or credit decisions. For instance, FRB-OCC (2011) specifies that reports generated from model outputs should be reviewed as part of the model validation process to ensure that they are accurate, complete and informative, and that they contain appropriate indicators of model performance and limitations. For the use of machine learning models for regulatory capital purposes, EBA (2021) recommends that banking institutions document the outcomes of statistical analyses involving risk drivers and output variables. The expectation for supported decision-making

[58] The report scored 10 major foundation developers based on 100 transparency indicators and found that the average score was only 37, with the top score being 54 out of 100.

becomes more explicit when adverse actions are taken. The US Consumer Financial Protection Bureau (CFPB) (2022), for instance, mandates creditors to provide applicants with specific reasons when an adverse action is taken against them. In some cases, the obligation to offer an explanation only arises if the customer requests it.

The principle around providing understandable explanations is broadly reflected in AI guidance and MRM frameworks, including the need to tailor them to specific audiences. European Commission (2019) underscores that when an AI system significantly impacts people's lives, stakeholders should be able to request a suitable explanation of its decisions. This explanation should be timely and tailored to the expertise of the specific stakeholders, whether they are consumers, regulators, or internal auditors. In the case of consumers, the right to be informed immediately and free of charge is contained in EU consumer credit law.① Moreover, the EU AI Act grants individuals the right to obtain clear and meaningful explanations from deployers regarding the role of the AI system in the decision-making process and key factors influencing the final decision. FRB-OCC (2011) stress that reports should account for the fact that decision-makers and model developers often come from different backgrounds and may interpret the same information differently. EIOPA (2021) notes that while simplified explanations are essential for non-technical stakeholders, such as consumers, technical stakeholders – like auditors – require more detailed and comprehensive information to effectively carry out their responsibilities.

Regarding expectations on explanation accuracy, AI guidance and MRM frameworks generally expect financial institutions to provide accurate and adequate explanations. There is a growing consensus around the need to disclose material information about AI-driven decisions. Information is considered material if its omission could influence stakeholders' decisions. To reduce subjectivity in determining materiality or adequacy, the Monetary Authority of Singapore (MAS) (2018) and the Hong Kong Monetary Authority (HKMA) (2024a) have specified that financial institutions must inform data subjects about their use of AI, associated risks, and how customer data is being used. Moreover, HKMA (2024a) suggests that financial institutions should disclose the factors influencing AI-driven decisions.

With respect to the expectations around communicating or understanding the capability limits of AI systems, AI guidance generally requires firms to communicate their capabilities, limitations and risks to relevant stakeholders. For instance, EIOPA (2021) stresses the importance of highlighting system limitations. The Central Bank of Brazil underscores that the board and senior management should have a clear understanding of the limitations and uncertainties involved in risk assessments, particularly when models are developed by third-party vendors. In this regard, the UK Prudential Regulation Authority requires vendors to provide appropriate testing results showing that their systems operate as expected, and to clearly indicate the circumstances in which the systems' use may be problematic. To address the limitations of AI systems, FRB-OCC (2011) recommend mitigating model uncertainty by incorporating human judgment, reducing reliance on the model's output, or ensuring that the model is supplemented by other models or approaches to more effectively manage associated risks. In the case of Qatar, the central bank's AI Guideline specifies that an entity must ensure that the human overseer is given tools and authority to intervene in the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure (Qatar Central Bank (2024)).

① See European Parliament (2023), Article 11.4.g.

Governance and accountability

29. **Transparency leads to greater accountability.** Accountability relates to having clear roles and responsibilities, as well as assigning ultimate responsibility to the board and senior management of a financial institution. Transparency, for example through the documentation of how the AI model works and the control processes surrounding it, makes assessing the fulfilment of these responsibilities much easier. AI policies typically accentuate the importance of traceability by maintaining documents or information before and after model deployment, with an appropriate retention period. Key elements to be documented include model changes and audit logs (who did what, and when);[59] preliminary

[59] See EBA (2020).

assessments;[60] usage details (such as databases accessed and if data matched verified identities);[61] trails to support AI system outcomes; project documentation; various versions of the model code; and the original data set used to develop, retrain and recalibrate the model.[62] Development of new AI applications is becoming quicker, but the time needed to assess and validate those models typically requires longer timeframes due to firms' internal accountability processes.

30. **To ensure greater accountability, AI guidance emphasises the role of human intervention.** This is to minimise the risk that AI-based decisions result in harmful outcomes, especially if the AI outputs have significant potential impact on customers. Hence, concepts like "human-in-the-loop" (human intervention in the decision cycle of the AI), "human-on-the-loop" (human intervention during the design cycle and subsequent reviews) and, more recently, "human-in-control" (primacy of humans in making critical decisions) are emphasised.

Reliability/soundness

31. **Expectations regarding reliability/soundness of AI models are closest to those for traditional models.** These involve the usual regular independent testing or monitoring to confirm that a model is performing as intended. They include monitoring metrics on validity, accuracy, robustness and reliability. What seems to be different is that ensuring reliability/soundness is viewed from the perspective of avoiding causing harm to customers due to decisions based on inaccurate decisions or inappropriate advice. As such, AI risk management efforts are expected to prioritise the minimisation of potential negative impact and emphasise the role of human intervention in cases where AI models cannot detect or correct errors.[63]

Fairness, ethics and safety

32. **Fairness is generally highlighted in the context of AI use in finance.** Two dimensions of fairness are mentioned in AI guidance: distributive fairness and procedural fairness. Distributive fairness relates to the fairness of outcomes resulting from AI-driven decisions; that is, AI should be non-discriminatory. This is the most often cited dimension of fairness in regulatory guidance. However, it is also the most challenging to measure and achieve. There are three major categories of AI bias – systemic; computational and statistical; and human-cognitive – and each can occur even in the absence of prejudice, partiality or discriminatory intent.[64] Procedural fairness, on the other hand, relates to the fairness of the decision-making process. The concepts of external transparency and external accountability, therefore, operationalise procedural fairness. While theoretically easier to achieve than distributive fairness, disclosures to customers about how an AI model works and how it came up with a decision could prove challenging. This issue is especially acute when it comes to gen AI.

60 See Federal Senate, Brazil (2023).

61 EU AI Act requirement for high-risk AI systems.

62 Qatar Central Bank (2024).

63 NIST (2023).

64 NIST (2023) describes the three major categories of AI bias as follows: *"Systemic bias can be present in AI datasets, the organizational norms, practices, and processes across the AI lifecycle, and the broader society that uses AI systems. Computational and statistical biases can be present in AI datasets and algorithmic processes, and often stem from systematic errors due to non-representative samples. Human-cognitive biases relate to how an individual or group perceives AI system information to make a decision or fill in missing information, or how humans think about purposes and functions of an AI system. Human-cognitive biases are omnipresent in decision-making processes across the AI lifecycle and system use, including the design, implementation, operation, and maintenance of AI."*

33. **Ethics is now somewhat folded into AI governance and expectations on accountability.** Ethics is broader than fairness issues and covers privacy and data protection, non-discrimination and equality, diversity, inclusion and social justice. It is based on a society's norms or mores, which may be codified in laws, regulations, codes of conduct, etc. To enforce this aspect, some regulatory guidance imposes a number of measures, including establishing an ethical code of conduct on the use of AI; putting in place policies for the procurement and lawful processing of data; seeking diversity in the input data; and carefully reviewing training and validation data during the model training process.

34. In terms of safety expectations, many AI guidance documents emphasise that AI systems need to be used in a way that avoids causing harm or infringing on human rights. This guidance requires that societal values, including fairness and ethical standards, be integral to the design, development and deployment of AI systems. To achieve this, the guidance refers to continuous monitoring and human oversight as necessary to ensure that AI systems operate as intended. Moreover, it highlights the importance of developing effective labelling and content provenance mechanisms to determine when content has used AI. While some jurisdictions have set up specific bodies to oversee compliance with AI safety standards (eg the UK and US AI Safety Institutes), others deal with this issue as part of their broader oversight of AI (eg the European AI Office, which includes a safety unit) or in the context of online safety research (eg the Singapore Centre for Advanced Technologies in Online Safety). IAIS (2024a) highlights the importance of insurers taking steps to observe existing legal requirements, including anti-discriminatory requirements, when adopting AI systems.

Data privacy and security

35. **Data privacy/protection and safety as well as security have become more prominent in newer AI guidance.** The importance of large quantities of data for delivering reliable/sound AI outcomes, coupled with fairness and ethical expectations for AI systems' design and operation, have enhanced policymakers' attention to safeguarding personal data such as individuals' identities, locations and habits. Additionally, AI systems can be used to mislead and manipulate individuals through, for instance, deepfakes and psychological profiling, resulting in complex and increasingly convincing forms of fraud and disinformation. This makes it crucial to develop and operate safe AI systems, ie aligned with societal values. Finally, growing reliance of businesses on AI systems and their increased exposure to cyber attacks and other malicious actors' attempts to exploit weaknesses makes it indispensable to deploy secure AI systems that are able to continue providing products and services despite disruptions.

36. **The right of individuals to data privacy/protection is emphasised, particularly when their personal information is at stake.** Accordingly, in line with applicable data-related laws and regulations, AI guidance requires individuals' consent for the collection, use and retention of personal data. These data should be safeguarded from privacy and confidentiality risks. AI providers are also expected to effectively respond to individuals' requests for, among others, data correction, supplementation and deletion. The EU guidance goes further and requires a strict process for detecting and correcting biases involving special categories of personal data, eg racial/ethnic origin, religious beliefs, health/biometric data and sexual life/orientation. The emergence of gen AI has increased attention to the personal data aspects of AI. For example, the draft guidance in China expects providers to comply with relevant data privacy laws and regulations as part of the entire process of training data used in AI systems.

37. **AI systems are expected to rely on sound security and resiliency standards.** Secured AI systems are those that can maintain their confidentiality, integrity and availability in the event of a disruption, including serious cyber security breaches. To achieve this, AI guidance generally outlines organisational and technical expectations for AI systems, including third-party risk management, typically following a risk-based approach. For instance, a high-risk system is expected to operate under a strong control environment and cyber security framework that prevents unauthorised employees and third parties from exploiting potential vulnerabilities. That said, if a serious cyber-related incident were to happen, AI

guidance (eg, Brazil[65], China[66] and the EU[67]) increasingly envisages reporting or communication to the competent authority and backup plans to promptly resume disrupted AI-related services.

Consumer redress and AI literacy/awareness

38. **The external dimension of accountability, including the requirement for consumer redress, is also often highlighted.** This reinforces the expectations on external transparency. Aside from the information described above that should be disclosed to customers, financial institutions using AI that may have a significant potential impact on customers should provide them with channels to inquire about, submit appeals for, and request reviews of AI-driven decisions that affect them. For instance, the EU AI Act and MAS (2018) envisage deployers of AI systems having mechanisms in place to take into account verified and relevant supplementary data provided by customers when performing reviews of AI-driven decisions.[68]

39. As gen AI becomes more integrated into everyday life, AI regimes seek to further improve AI literacy and awareness as well as to facilitate consumer redress. AI providers and deployers are increasingly expected to adopt awareness and/or training measures for their staff, including those involved in the operation and use of AI systems, as well as for individuals affected by AI systems, with special attention to vulnerable groups.

Other policy themes

40. **With the emergence of gen AI, many AI guidance documents are paying increased attention to intellectual property and sustainability considerations.** AI providers are expected to ensure compliance of gen AI systems with intellectual property laws. These mainly include obtaining appropriate licences or permissions for the use of training data; giving proper attribution to the original creators of copyrighted material; and explaining in a transparent manner how AI systems handle copyrighted content. In addition, given that gen AI systems require high-performance computing capabilities and hence large amounts of energy, these systems are expected to be developed and operate using standards for increasing energy efficiency. To help assess whether these expectations are met, AI providers are generally expected to keep records of relevant information related to AI system development, testing and operation. Their climate-related disclosure commitment may oblige them to disclose their carbon footprint arising from their AI-related services.

41. **Many AI guidance documents highlight the importance of international interoperability of AI guidance and public-private partnerships.** AI guidance includes references to the need to engage with the international community to support AI interoperability across different regulatory regimes, minimise cross-border frictions and facilitate local firms' compliance if they were to operate abroad. AI guidance also encourages public-private sector partnerships. It is increasingly envisaging strong collaboration between government, industry, academia and various representatives from civil society to ensure that AI systems can be effective in driving innovation while being developed and deployed in a responsible manner.

[65] See Section 38 of Federal Senate, Brazil (2023).

[66] See Article 43 of China's Draft AI Law (see CSET (2024).

[67] See Article 73 of the EU AI Act.

[68] See Article 18(8) in EU Parliament (2023) and Section 7 in MAS (2018).

Section 4 – Practical issues in implementing cross-sectoral AI guidance to the financial sector: the case of credit and insurance underwriting

42. **The common themes of cross-sectoral AI-specific guidance outlined in Section 3 are not new to the financial sector and hence are addressed through general financial regulations.** In the financial sector, these themes are addressed through general regulations covering governance, risk management (including model risk management, third-party risk management, operational risk/resilience and cyber security) and consumer protection. For a long time now, financial authorities have focused on making sure financial institutions have appropriate risk management and controls in place in running their businesses. This focus contrasts with the compliance-based approach of the past, where financial institutions needed to observe strict compliance with detailed rules. This recognises that the role of financial authorities is not to manage financial institutions but to ensure that they operate in a safe and sound manner at all times. This extends to the technologies, including AI, that financial institutions are using to run their businesses.[69]

43. **Not many financial authorities have issued regulations specifically addressing financial institutions' use of AI.** Financial authorities have so far issued high-level principles (eg EBA, EIOPA, HKMA, MAS) or clarification as to how existing regulations apply to AI (eg UK authorities). So far, among the authorities examined for this paper, only the Qatar Central Bank (QCB) and several US state insurance regulators[70] have issued regulations specifically addressing AI use by financial institutions. The regulations contain specific rules that regulated entities need to follow when developing, purchasing and deploying AI systems, or when outsourcing processes or functions that rely on AI. The EBA and EIOPA may follow suit in order to clarify the relevant provisions of the EU AI Act, which classifies evaluating creditworthiness and risk assessment and pricing of health and life insurance as being among the high-risk uses of AI systems (see Box 3). These use cases are in the areas of credit and insurance underwriting. In the insurance sector, the IAIS has examined its Insurance Core Principles (ICPs) and concluded that they are sufficiently principles-based to capture AI risks. IAIS (2024a), when finalised, will provide a clear framework, consistent with the ICPs, for addressing risks that insurers face when using AI systems.

Box 3

Risk-based policy approaches and high-risk AI systems

AI guidance appears to increasingly follow a risk-based policy approach to deal with AI systems (eg the EU's AI Act; Brazilian Draft Bill 2338/2023 on AI; Qatar Central Bank – QCB AI Guideline). This approach is generally designed to address the potential harmful effects of AI systems on fundamental human rights and democratic values. The greater this potential harmful effect, the more stringent restrictions are imposed by policy frameworks, including prohibiting some AI-related activities.

Based on this criterion, the Brazilian Draft Bill classifies AI-related risks into excessive, high and other risks. Along the same lines, the EU AI Act uses a similar but more granular classification: unacceptable, high, limited and minimal/no AI-related risks. In both policy frameworks, when AI activities are categorised as generating excessive or unacceptable AI-related risks, these are prohibited. The EU AI Act provides examples of prohibited AI-related activities including social scoring systems, subliminal behavioural manipulation and real-time biometric identification in public places for law enforcement.①

69 More recently, however, some financial authorities have issued technology-related regulations (ie cloud-specific regulations) to address heightened security risks that cloud use brings. In general, however, cloud use is still covered under general IT risk management, operational risk, operational resilience and third-party risk management regulations.

70 National Association of Insurance Commissioners (2023) is a model bulletin that US state insurance regulators can use to enact laws or issue guidelines on the use of AI by insurers. Several US states have issued insurance-specific AI regulations or guidance – see here.

Another similarity across risk-based policy approaches is that most of these frameworks are largely centred on high-risk AI systems. Focusing on finance, the Brazilian Draft Bill considers high-risk AI systems when these are used for assessing the debt capacity of individuals, establishing credit ratings or biometric identification.② Following a similar approach, the EU AI Act considers as high-risk AI systems those that are used to evaluate the creditworthiness of natural persons or establish their credit score. Additionally, in the EU, AI systems are considered high-risk when used to undertake risk assessment and pricing in relation to natural persons in the case of life and health insurance. The QCB AI Guideline defines high-risk AI systems as those that have the potential to cause a significant negative impact on an entity's operations or the financial system.③

In the EU AI Act, different requirements are imposed on providers and deployers of high-risk AI systems. Requirements imposed on the former are more stringent and include those relating to risk management systems, data governance, technical documentation, record-keeping, transparency and provision of information to deployers, human oversight, accuracy, robustness and cyber security. Deployers of high-risk AI systems, on the other hand, must ensure that they use the AI system in accordance with the instructions for use, apply suitable human oversight, monitor and keep logs of its operation, and inform workers' representatives when using that technology in the workplace.

① See Article 5 in European Parliament (2024). ② See Article 17 in Federal Senate, Brazil (2023). ③ See Section 2, definition 10 in Qatar Central Bank (2024).

44. **Underwriting is a core process of lenders and insurers that is likely to become a focus for AI regulatory work by financial authorities.** In general, it is a process by which a financial institution determines whether an applicant is qualified to be granted a financial product (ie loan or insurance) and at what price. In credit underwriting, the lender assesses the probability that an applicant can repay the loan. This involves reviewing an applicant's capacity and willingness to pay by looking at factors such as credit history, income, employment stability and other liabilities. In insurance underwriting, the insurer assesses the relevant risk of the applicant to determine the appropriate level of premium to charge. For life insurance, for example, this entails gathering information on an applicant's medical history, lifestyle, age, etc. In both cases, sound underwriting practices can minimise losses either from too many defaults or insufficient premiums to cover claims. As discussed in Section 2, AI has the potential not only to address and mitigate some of the challenges facing financial institutions in credit and insurance underwriting but also to further enhance their capabilities in this area. Since this is a core financial and economic activity, it is likely that the use of AI in underwriting will attract the attention of financial authorities. Discussions with authorities for this paper suggest the following areas may be particularly relevant:

- governance and accountability;
- transparency and explainability;
- use of third-party AI services, data security and operational resilience; and
- new players and new business arrangements.

Governance and accountability

45. **Expectations with regard to governance and accountability outlined in cross-sectoral AI-specific guidance are very similar to those already required for financial institutions, including in the conduct of underwriting activities.** More specific accountabilities for underwriting include: (i) model owner – this individual or team holds overall responsibility for the development, implementation and use of the underwriting models; (ii) model developers – these are tasked with developing, testing, evaluating and documenting the underwriting models; and (iii) model users: typically, these are credit officers or insurance underwriters who rely on the model's output to inform underwriting decisions. Each of these tasks are expected to align with the firm's AI risk management framework and risk appetite.

46. **It is not surprising, therefore, that newly developed AI risk management frameworks reference the general governance principles.** For example, ISO/IEC 23894[71] relies on its existing risk management standards (ISO 31000:2018). The NIST AI Risk Management Framework[72], on the other hand, is based on four all-too-familiar functions: govern, map, measure and manage. Nevertheless, both standards also emphasise the unique considerations relating to AI. For example, privacy risk, fairness and bias are specifically highlighted in the NIST standards, as well as the role of human oversight. Its gen AI companion resource, meanwhile, draws out risks amplified by the technology, such as those related to information integrity and intellectual property. In terms of human oversight, it should be noted that there is a trade-off between human intervention requirements versus the intended operational efficiency objectives when firms use AI. Autonomous AI systems that can make their own decisions, eg automated acceptance of loan or insurance applications, could be seen as contradicting the human intervention requirements.

47. **Applying these governance principles in the context of AI will require the necessary expertise and skills.** Financial institutions' board and senior management will need to have a sufficient level of AI expertise or familiarity to be able to effectively carry out their governance responsibilities, such as providing effective challenge to AI-driven decisions and assessing their broader impact on the institution's business strategy. Similarly, financial institutions' staff will need to have the requisite skills to effectively develop, deploy and manage the risks from using AI systems, as well as provide independent internal assurance. More concretely, financial institutions face the challenge of ensuring that they have the necessary expertise to develop or maintain AI systems that are not only high-performing but also comprehensible to internal stakeholders (eg board of directors and senior management) and viewed as fair and reliable by external stakeholders (eg clients and regulators). As use of AI by financial institutions increases, financial authorities will also need similar skills to be able to effectively regulate and supervise.

Transparency and explainability

48. **The implementation of these governance principles will also be affected by the issue of AI explainability.** As mentioned, explainability refers to making transparent how an AI system's outputs (eg underwriting decisions) were derived from its inputs (eg customer data). This includes providing clarity as to how the system functions and makes decisions. However, as AI systems become more complex, they often achieve higher performance at the expense of explainability. In other words, while these systems can leverage large, diverse sources of credit- or insurance-related information and detect intricate data patterns, this increased complexity can make their decision-making processes harder to understand. Therefore, striking the right balance between performance and explainability is one of the main challenges for financial institutions implementing AI, especially in credit and insurance underwriting. Due to these explainability challenges, some industry players advocate that regulations should focus on the risk control surrounding the use of AI rather than on explainability or transparency metrics. Others are proposing to focus on AI outputs, ie placing emphasis on whether the decisions or predictions made by the AI are fair, ethical and compliant with regulations, regardless of how the AI arrives at these outcomes.

49. **The lack of transparency in how AI systems make credit and insurance decisions raises significant concerns about compliance with consumer protection and model risk management (MRM) requirements.** Consumer protection regulations generally require financial institutions to inform clients of the primary reasons behind credit or insurance application denials, under the so-called "adverse action" requirements. Moreover, MRM frameworks are crucial tools for managing and mitigating AI-related risks, including issues of opacity and lack of explainability. Financial institutions are expected to

[71] See ISO (2023).

[72] See NIST (2023).

address these risks as part of their evaluation of model complexity. This often requires enhancing oversight of AI models, with particular attention to validation processes and risk control measures.[73]

50. **For insurance underwriting specifically, policyholder protection is a key objective of insurance regulators, be they prudential or conduct regulators.** As such, the issue of unfair treatment of customers that could arise from the use of AI in insurance underwriting attracts great regulatory scrutiny. New York Department of Financial Services (2024a) provides detailed guidelines in relation to governance and risk management, fairness and transparency for the use of AI in insurance underwriting and pricing.[74] The overarching fair treatment rules apply to the entire insurance underwriting process, from ensuring the data inputs are not biased and that data privacy laws are respected, to providing proper customer disclosure before concluding a transaction. Some life insurance products with savings or investment elements require extensive customer due diligence process. The use of AI to underwrite such products may be challenging, as the system will need to understand the context of the applicant before recommending the insurance/financial product. IAIS (2024a) called for insurance supervisors to ascertain that insurers are able to meaningfully explain the outcomes of AI systems, covering how decisions or predictions are made especially for use cases that could have a material impact on solvency or consumers.

51. **The transparency expectation, including its consumer protection aspect, and its interconnection with fairness and ethics expectations emphasise the role of data governance and data management.** AI systems need to be properly documented, including their design, processes and data used. Documentation of data used is particularly important to be able to explain AI-based outcomes or decisions to customers, and in assessing which supplementary data that may be provided by customers are relevant. Moreover, financial institutions need to assess whether data inputs are biased and put in place policies and measures to ensure that they are lawfully, ethically and securely collecting, storing, processing/using and sharing data (see below discussion on data security and privacy concerns arising from use of third-party AI). These factors point to the need for financial institutions to have robust data governance frameworks, as well as appropriate data management tools and procedures to enforce these frameworks.

52. **Use of gen AI in credit and insurance underwriting will further exacerbate explainability challenges.** These challenges stem largely from the complexity of how gen AI systems operate. These systems rely on billions or even hundreds of billions of parameters, making it difficult to trace how specific inputs lead to specific outputs and to understand the systems' internal decision-making process. Unlike traditional AI systems, where the same input always gives the same result, gen AI systems can give different results from the same input because they are designed to introduce an element of variability, which makes them flexible and adaptable but less deterministic. Additionally, since gen AI systems can create novel content, it becomes harder to explain the decisions behind these outputs. Finally, unpacking how a system might generate biased or ethically questionable content involves analysing intricate patterns in training data, which often requires highly technical approaches and may involve disclosing sensitive information.

53. **Various techniques are being explored to address concerns related to AI explainability in the credit context.**[75] For instance, some US financial institutions are tackling these issues by imposing upfront constraints on model complexity, applying post hoc techniques, or using a combination of both approaches.[76] Post hoc techniques aim to provide insights into how a model works or why it made a specific decision after it has already been trained. Examples of post hoc techniques include building surrogate models (SMs) and applying feature-importance techniques (FTs). SMs are simplified models that

[73] See, for example, BoE-PRA (2023).

[74] The guidelines prohibit insurers from using AI in underwriting or pricing unless they can demonstrate that they do not unfairly or unlawfully discriminate against consumers. The guidelines provide detailed steps that insurers need to undertake to make this assessment, including quantitative metrics that should be considered.

[75] The OECD provides a catalogue of tools and metrics to assess AI models.

[76] See FinRegLab (2021).

approximate how complex AI models make decisions, either across the entire data set or for individual consumers.[77] FTs explain a model's behaviour by quantifying the contribution of each input to a specific prediction (eg Shapley Additive Explanations (SHAP)).[78]

54. **While recent advancements in explainability techniques are promising, further work is still necessary.** Empirical analysis of machine learning models used in credit underwriting, including some complex models, indicates that not all explainability techniques reliably capture key aspects of model behaviour.[79] Additionally, the outputs of these techniques must be interpreted with a clear understanding of the underlying data used in credit underwriting decisions. This reflects the absence of a "one size fits all" explainability solution that works for all AI models. Supporting this, a recent EBA survey revealed the range of explainability measures employed by European financial institutions: Shapley values (40% of respondents), graphical tools (20%), enhanced reporting and documentation of the model methodology (28%) and sensitivity analysis (8%).[80]

55. **Financial authorities can play a role in promoting the consistent application of sound explainability techniques in AI-driven credit and insurance underwriting.** As a useful first step, authorities could define basic concepts and provide guidance on the key qualities to consider when selecting explainability techniques and assessing their effectiveness. This regulatory intervention by outlining key criteria and expectations can be helpful in accelerating improvements and fostering consistent implementation of sound explainability techniques across the financial industry. Incorporating these features into MRM frameworks would provide a practical foundation for further progress. In addition, consumer protection regulations may need to be refreshed to clearly articulate the types of disclosures required when individuals are denied credit or insurance based on AI decisions.

Use of third-party AI services, data security and operational resilience

56. **Use of third-party AI services[81] by financial institutions appears to be prevalent and increasing, which poses another challenge.** While there is no authoritative source of data on the use of third-party AI services by financial institutions, there are different sources of information that, when combined, may give a good indication. For example, a 2023 cross-sectoral survey of 1,240 respondents representing business organisations – including financial institutions – in 87 jurisdictions revealed that 78% of the respondents were using third-party AI models, with 53% using exclusively such models.[82] Among financial institutions, the majority expected that use of third-party AI models would increase by 10–25% in the next 12 months.[83] For credit modelling specifically, a survey of small to mid-sized financial institutions in the United States showed that 20% did not have in-house staff for credit modelling and

[77] An example of local SMs is LIME – Local Interpretable Model-agnostic Explanations.

[78] In the insurance sector, professional actuarial bodies have issued model risk management guidelines which cover AI models. For example, Financial Reporting Council (2024) provides guidance on model governance, how to identify material biases and limitations of models. It includes a case study on how to communicate the performance of AI models to a non-technical audience. Actuarial Association of Europe (2024) describes approaches to AI explainability including LIME and SHAP.

[79] See FinRegLab (2021).

[80] See EBA (2023).

[81] Broad examples of AI services that third parties may provide to financial institutions include: (i) providing the AI model itself that financial institutions then customise to their use; (ii) processing data from financial institutions using AI models, with the processed data becoming input to financial institutions' own models; and (iii) providing output of AI models to financial institutions, which in turn use it as input to their own models (see, for example, Veritas Initiative (2023)).

[82] See MIT-BCG (2023).

[83] See IIF-EY (2023).

outsource this function to a third party.[84] So the extent of use of third-party AI services by financial institutions appears significant and financial authorities need to examine and address its potential risks.

57. **The regulatory principle that financial firms' board and senior management is ultimately accountable for any activities, functions, products or services provided by third parties also applies to AI.[85]** For example, at a high level, financial institutions are expected to have appropriate processes in place for selecting third-party AI models and making sure that these are validated to the same standards as their own internally developed models. To this end, contracts or agreements between financial institutions and third parties are expected to include clauses requiring third parties to provide evidence that the model is appropriate for the financial institution's intended use; testing results that show the model works as expected; and information on the model's limitations and assumptions. Third parties are also typically expected to conduct ongoing performance monitoring and outcomes analysis and make appropriate modifications over time.[86] In some jurisdictions, contractual clauses providing supervisory authorities the right to audit third parties are also included.

58. **While this guiding principle is sound, in practice and in the context of AI, it can be challenging.** Third-party models may not allow financial institutions full visibility of certain proprietary information, eg the computer coding and other details. Requiring disclosure of such information could expose third parties' intellectual property and confidential business information. This, in turn, could disincentivise innovation and further AI development. Hence, it is recognised in regulations that in some cases financial institutions may need to modify their approach. For example, when validating third-party models, financial institutions may need to rely more on sensitivity analysis and benchmarking.[87]

59. **One proposal to address this challenge is to clearly delineate the responsibilities of users of AI services (ie financial institutions) and their providers (ie third parties) based on what each can control.** This is the approach being advocated by technology firms providing AI services and borrows from the shared responsibility model for cloud computing services.[88] For example, third parties that provide AI models to financial institutions have control over the development of the base/foundation AI model and should thus be responsible for providing documentation in this regard. Financial institutions, on the other hand, have control over how the third-party AI model is deployed and retrained; thus regulators can look to them to ensure that related processes are sound.

60. **In the context of credit and insurance underwriting, the remaining question is whether this arrangement is enough to meet the policy expectations outlined in Section 3.** In terms of assessing reliability or soundness of the model, financial institutions' ongoing monitoring and analysis of third-party model performance using outcomes from financial institutions' own use could be sufficient. Achieving procedural fairness (ie external transparency and accountability), however, may still pose a challenge. It is not clear whether financial institutions would be able to adequately explain to customers AI-driven decisions that are largely influenced by foundation models rather than by the customisation that they have done. Moreover, financial institutions almost certainly would face heightened reputational risk. Even if third parties would be required to make appropriate disclosures on their foundation models, data or assumptions, if something were to go wrong, it would be likely that financial institutions would be blamed by customers regardless of whether they built or bought the AI model. In any case, requiring third parties to disclose to customers factors within their control that affect AI-driven decisions implies that third parties need to be identified and be subject to oversight by financial authorities.

84 See Cornerstone Advisors (2020).

85 See IAIS (2024a).

86 See, for example, FRB-OCC (2011).

87 Ibid.

88 See Veritas Initiative (2023).

61. **Use of third-party AI for credit and insurance underwriting raises data security and privacy concerns.** AI systems that handle sensitive and personal customer data – such as those used for credit and insurance underwriting – are attractive targets for cyber attacks, data breaches and abuse. They could also be subject to data poisoning attacks, which attempt to corrupt and contaminate training data to compromise the system's performance. These highlight the need to manage the risks of sharing data with third parties. This could be done, for example, through master service agreements that set out requirements relating to data maintenance, access, rights, ownership and intellectual property, and security requirements. Financial institutions could also conduct due diligence on third parties to assess their data controls and ethical reviews on how the third party will use the data.[89]

62. **Use of third-party AI services – and its relationship with cloud services – presents operational resilience issues.** Use of third-party AI services (eg data processing and provision of AI model output) may be facilitated through APIs. Financial institutions are also increasingly moving their core business workloads – including credit and insurance underwriting – to the cloud.[90] In addition, the large providers of AI services are also the major cloud service providers (CSPs), which deploy their AI services through their cloud infrastructure. All these factors result in more interconnectivity that also makes financial institutions more vulnerable to cyber threats and operational disruptions at AI service providers.[91]

63. **Given the close link between cloud and AI services, the need for a more direct approach to the oversight of third parties to safeguard operational resilience is becoming stronger.** Currently, financial authorities typically follow an indirect approach in addressing operational resilience issues resulting from third-party services, including AI. This approach relies on financial institutions to manage the risks from third-party services and to assess the potential implications of such services for their own operational resilience. For example, financial institutions are required to verify that third parties have at least an equivalent level of operational resilience to that expected by financial authorities. However, financial institutions might not have full visibility into the risk management and control measures adopted by third parties. In addition, while the indirect approach could potentially address risks faced by individual financial institutions, it may not be sufficient to address the potential impact on the financial system of an operational disruption of a third party that provides services to multiple financial institutions.[92] Hence, a few jurisdictions now have or are planning to have direct oversight by financial authorities over third parties that are considered critical to the functioning of the financial system. As more financial institutions use cloud and AI services provided by the same third parties, some jurisdictions may find there is increasingly a clear case for having a more direct oversight approach for these third parties.

New players and new business arrangements

64. **Ensuring that regulatory expectations relating to the use of AI are also met by non-bank lenders is another challenge.** This is especially the case when it comes to new entrants, such as fintech and big tech lenders. These lenders use digital delivery channels and rely on alternative data for credit underwriting. Moreover, non-bank lenders with digital business models are said to be more established users of AI models.[93] In many cases, these lenders may be subject to different sets of regulations from bank lenders. This may be justified by the fact that their activities pose different risks to those of traditional players. In any case, it may be prudent to examine regulations relevant to these players to determine if they require adjustments to take account of the cross-sectoral expectations on the use of AI. This would help avoid regulatory gaps in addressing risks arising from the use of this technology. The case of big tech

89 See BCBS (2024).

90 See Koh and Prenio (2023).

91 See IAIS (2023a).

92 See Prenio and Restoy (2022).

93 See FinRegLab (2021).

lenders is especially interesting. Some of these have significant lending activities,[94] while at the same time they may be providing cloud and AI-related services to banks and other lenders. The risks they pose, therefore, span various aspects of the banking value chain.

65. **Novel arrangements in delivering lending and insurance products to customers, such as through bank/insurer partnerships with fintech or big tech firms, further complicate the enforcement of regulatory expectations.** Banking-as-a-Service (BaaS), for example, allows banks to provide credit through non-bank intermediaries (eg fintech/big tech firms and other non-financial firms) that serve as interfaces to clients.[95] In the case of non-bank intermediaries, this arrangement increases the use and value of their digital platforms by offering banking products while remaining outside the regulatory perimeter. In the case of banks, this arrangement enables them to access new customers and leverage the non-bank intermediaries' technological capability. In this type of arrangement, banks typically make the credit decisions, but the customer relationship is with the non-bank intermediaries.[96] In the insurance sector, big techs may serve as insurance intermediaries through embedded insurance or insurance marketplaces. They may also act as providers of technology services (eg cloud computing) or data services.[97] It is therefore unclear who should be responsible for ensuring that regulatory expectations regarding external transparency and accountability are met. This is further complicated if the AI models used by banks in driving credit decisions are provided by third parties. In general, as these multi-layer arrangements become more prevalent in the financial system, enforcing regulatory expectations on the use of AI could be a challenge.

66. **Understanding and addressing these practical issues is important for the safe and responsible adoption of AI by financial institutions.** Some financial authorities are already actively working with the industry to achieve this. Together with the industry, the MAS has co-created the Veritas Initiative, which aims to enable financial institutions to evaluate their AI solutions against the MAS FEAT Principles[98]. The Veritas Initiative developed the FEAT assessment methodology and has tested integrating the methodology into financial institutions' existing governance frameworks as well as specific use cases. The HKMA, on the other hand, recently launched its GenA.I. Sandbox[99], which aims to promote responsible innovation in gen AI across the banking industry. The Sandbox provides a platform for banks to pilot their gen AI use cases within a risk-managed framework, supported by essential technical assistance and targeted supervisory feedback.

Section 5 – Conclusion

67. **The broader adoption of AI has the potential to bring transformative benefits to society as a whole and to the financial system in particular.** Within the financial system, AI capabilities offer opportunities to financial institutions to substantially enhance productivity as well as to achieve time and cost efficiencies in their activities. AI also offers unprecedented levels of automation and accuracy in regulatory compliance, including fraud detection and AML/CFT. By analysing vast amounts of structured and particularly unstructured data, AI holds the promise of enhancing customer experiences and contributing to a more inclusive financial system.

94 See Cornelli et al (2023).

95 See BCBS (2024b).

96 See Barakova et al (2024).

97 See Garcia Ocampo et al (2023).

98 See MAS (2018).

99 See Press Release.

68. **The use of AI by financial institutions – while potentially exacerbating existing risks – currently does not appear to present new ones.** Use of AI may have negative consequences for equality, privacy and the environment, among other factors. Given these significant societal implications, it is thus not surprising that governments around the world are coming up with legislation or regulations to ensure that AI is safely and responsibly used. However, examining the risks AI poses when used by financial institutions, one would come up with the usual list of risks that are already familiar to financial institutions and financial authorities. Admittedly, AI use may heighten some of these risks, such as model risk (eg lack of explainability makes it challenging to assess appropriateness of AI models) and data-related risks (eg privacy, security, bias). Financial institutions are therefore working to enhance their controls and tools to manage these risks, while financial authorities are building capacity to oversee them.

69. **Consequently, the common themes of cross-sectoral AI-specific guidance are already broadly covered in existing financial regulations, so the need for separate and comprehensive AI financial regulations could be arguable.** This is perhaps the reason why financial authorities in most jurisdictions are not planning to issue specific AI regulations in the near future. On the other hand, industry players may be waiting for greater clarity on regulatory stance before investing billions in developing AI applications that may be constrained or prohibited by future regulations. The proliferation of AI definitions also seems to underscore the challenge of capturing in words the essence of this evolving technology. It is hard to regulate something that is in flux. This is the reason why regulators are in general taking a technology-neutral approach. On the other hand, uncertainties created by overly wide definitions can inadvertently capture non-high risk AI systems that have been used by firms for decades. The pragmatic way forward, it seems, is to ensure that the desired regulatory outcomes are achieved regardless of what technologies financial institutions use.

70. **Nevertheless, AI presents some unique challenges in implementing existing financial regulations and hence AI-specific regulatory or supervisory guidance may be needed in certain areas.** This points to the need to examine existing regulations and, if necessary, consider issuing clarifications, revisions or even new regulations especially with respect to use cases that present higher risks or significant potential impact on customers. In particular, at least in the context of credit and insurance underwriting, the following areas stand out as important:

(i) **Governance framework.** The board and senior management of financial institutions are ultimately accountable for their activities, including AI use cases. That said, financial institutions' use of AI, particularly in core business activities, underscores the importance of a clear allocation of roles and responsibilities across the entire AI life cycle (ie design, delivery and deployment of AI). Governance frameworks might need to specify the role of human intervention to minimise harmful outcomes from AI systems.

(ii) **AI expertise and skills.** A foundational element to effectively implementing, managing and overseeing AI systems is having the necessary expertise and skills that may not be widely available currently in financial institutions, including at the board and senior management level. The type of expertise and skills needed would partly depend on the regulatory/supervisory approach to AI and the principles of proportionality.[100] Moving forward with a wider adoption of AI without the corresponding expertise and skills could result in insufficient understanding and ineffective management of the risks to financial institutions and the financial system. Financial authorities may therefore consider clarifying their expectations regarding the expertise and skills envisaged to be in place for financial institutions that plan on expanding AI use in their core business activities.

(iii) **Model risk management.** In the context of AI, and particularly gen AI, financial authorities may need to pay close attention to financial institutions' model risk management given the heightened

[100] Financial institutions are not expected to employ data scientists in order to fully understand LLMs for low-risk use cases. The skills required would also depend on, for example, the regulatory requirements relating to explainability.

model risk caused by, for example, lack of explainability of AI models. Some financial authorities already have model risk management regulations in place. Some have model risk management regulations that are specific to models used for regulatory purposes (eg calculating regulatory capital). Other authorities try to capture some elements of model risk management in general risk management regulations. In the first case, it might be helpful to define basic concepts and provide guidance on the key qualities to consider when selecting explainability techniques and assessing their effectiveness. In the last two cases, it might be worthwhile for financial authorities to consider issuing model risk management regulations that capture all types of models used by financial institutions, including AI.

(iv) **Data governance and management.** Considering increased data-related issues from the use of AI, financial authorities may also need to pay close attention to financial institutions' data governance and the data management tools and procedures that enforce it. Many of the relevant elements of data governance and management are captured in existing regulations, such as for model risk, consumer privacy and information security. Financial authorities may want to assess whether these are enough or need strengthening, or whether there is a need to issue regulations that address all data governance and management-related issues. Financial authorities can also support effective data governance and management by taking stock of the range of practices across financial institutions and promoting better practices.[101]

(v) **New/non-traditional players and new business models/arrangements.** To avoid potential regulatory gaps, regulations relevant to new/non-traditional players providing financial services would need to be assessed to determine whether they require adjustments to take account of the cross-sectoral expectations on the use of AI. A similar regulatory assessment might be needed with respect to multi-layer arrangements in providing financial services (eg BaaS) involving AI that may make it challenging for financial authorities to attribute accountability to various players in the ecosystem.

(vi) **Regulatory perimeter – third parties.** The concentration of cloud and AI service providers to a few large global technology firms strengthens the argument for putting in place direct oversight frameworks for these service providers.[102] In response, some jurisdictions have already moved in this direction, while others have reinforced the financial institutions' responsibility to manage risks stemming from these third-party relationships. This indirect approach is prevalent in the financial sector.

71. **Other areas not covered in this paper may be worth exploring in further research.** Examining the following areas may provide financial authorities with additional perspective on the implications of AI use by financial institutions:

(i) **Risk management of financial institutions.** Many papers looking at AI use in finance focus on the investments made by financial institutions in integrating AI capabilities into their businesses and operations. However, there is not much focus on the risk management spending of financial institutions to address heightened risks from AI use. Although it is reasonable to assume that the spend on risk management would not increase linearly with the increased spending on AI, some increase in budget allocation for risk management can be expected. Aside from spending, it would be worthwhile to study the actual risk management enhancements that financial institutions have introduced to identify, assess, address and mitigate risks arising from their AI-related activities. BCBS (2024) and IAIS (2024a) have outlined some of these risk management enhancements to address risks from gen AI. Further research can build on this and try to map heightened risks to enhancements in risk management practices.

[101] See BCBS (2024).

[102] Some insurers have noted that these providers have significant market power.

(ii) **Use of AI for regulatory compliance (regtech).** Financial institutions have been using AI to support AML/CFT compliance as well as in calculating regulatory capital. In general, the use of AI for regulatory compliance – especially if the models are similar or provided by the same vendors – leads to concern about concentration and herding behaviour. In the two examples cited above, an error in the models could have financial integrity and financial stability implications. Further research can look at how AI is used for regtech purposes and the risks this poses to regulatory objectives.

(iii) **Supervisory approaches by financial authorities to oversee the use of AI.** Upskilling, acquiring and retaining AI expertise within financial authorities is imperative to be able to provide effective supervisory oversight in the area of AI. This expertise can also be helpful in allowing authorities to take fuller advantage of this technology in the delivery of their supervisory responsibilities (suptech). Moreover, financial authorities may have different approaches in categorising AI systems and in applying risk-based supervision. Further work to describe different approaches in these areas would be helpful.

72. **Collaboration among financial authorities both domestically and internationally is important in continuing to understand and monitor risks from AI as the technology evolves.** Collaboration, for example, could be used to have a better understanding of AI use cases in the financial sector. This would help identify the specific areas in the financial sector where there may be heightened risks. At the moment, data on AI use cases in finance are anecdotal at best. The presence of various definitions of AI across jurisdictions is a significant impediment to acquiring these data. Hence, international alignment of the definition is an obvious first step, while recognising that any agreed definition may have to be adjusted as the technology evolves. An agreed definition will facilitate the identification of risks and provide an idea of where they can be found.

References

Accenture (2024): "Banking in the age of generative AI", *Research Report*, February.

Actuarial Association of Europe (2024): "What should an actuary know about artificial intelligence?", *AAE Discussion Paper*, January.

Ahmed, N, M Wahed and N Thompson (2023): "The growing influence of industry in AI research", *Science Journal*, vol 379, no 6635, pp 884–86.

Aldasoro, I, S Doerr, L Gambacorta, S Notra, T Oliviero and D Whyte (2024): "Generative artificial intelligence and cyber security in central banking", *BIS Papers*, no 145, May.

Aldasoro, I, L Gambacorta, A Korinek, V Shreeti and M Stein (2024): "Intelligent financial system: how AI is transforming finance", *BIS Working Papers,* no 1194, June.

Bank for International Settlements (BIS) (2024): *Annual Economic Report*, June.

Bank of England (2022): "Artificial Intelligence and machine learning", *Discussion Papers*, no 5/22, October.

——— (2024): "Engaging with the machine: AI and financial stability", speech by Sarah Breeden at the HKMA-BIS Joint Conference on Opportunities and Challenges of Emerging Technologies in the Financial Ecosystem, 31 October.

Bank of England-Prudential Regulation Authority (BoE-PRA) (2023): "Model risk management principles for banks", *Supervisory Statements*, no SS1/23, May.

Bank of England (BOE) and Financial Conduct Authority (FCA) (2024): *Artificial intelligence in UK financial services – 2024*, November.

Barakova, I, J Ehrentraud and L Leposke (2024): "A two-sided affair: banks and tech firms in banking", *FSI Insights on policy implementation*, no 60, October.

Basel Committee on Banking Supervision (BCBS) (2022): *Newsletter on artificial intelligence and machine learning*, March.

——— (2024): "Digitalisation of finance", *BCBS Working Papers*, May.

Calabia, C (2024): *AI: transforming the future or triggering fear? Generative artificial intelligence and its impact on financial consumers and regulators*, March.

Center for Security and Emerging Technology (CSET) (2024): *Translation of the artificial intelligence law of the People's Republic of China*, May.

Consumer Financial Protection Bureau (CFPB) (2022): "Adverse action notification requirements in connection with credit decisions based on complex algorithms", *Consumer Financial Protection Circular 2022-03*, May.

Cornelli, G, J Frost, L Gambacorta, R Rau, R Wardrop and T Ziegler (2023): "Fintech and big tech credit: drivers of the growth of digital lending", *Journal of Banking & Finance*, vol 148:106742, March.

Cornerstone Advisors (2020): "Credit modelling and the need for speed: the case for advanced technologies", *Research Report*.

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

——— (2021): "EBA discussion paper on machine learning for IRB models", *EBA/DP/2021/04*, November.

——— (2023): "Machine learning for IRB models: follow-up report from the consultation on the discussion paper on machine learning for IRB models", *EBA/REP/2023/28*, August.

European Central Bank (ECB) (2024): *Financial Stability Review*, May.

European Commission (2019): *Ethics guidelines for trustworthy AI*, April.

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

——— (2024): *Report on the digitalisation of the European insurance sector*, April.

European Parliament (2023): "EU Directive 2023/2225 on credit agreements for consumers and repealing Directive 2008/48/EC", *Official Journal of the European Union*, October.

——— (2024): *Artificial Intelligence Act*, March.

Evident (2024): *Evident AI talent report*, July.

Federal Reserve Board and Office of the Comptroller of the Currency (FRB-OCC) (2011): "Supervisory guidance on model risk management", *SR Letter 11-7 Attachment*, April.

Federal Senate, Brazil (2023): "Bill no 2338, of 2023", May.

Federal Trade Commission (2024): "FTC launches inquiry into generative AI investments and partnerships", 25 January.

Financial Action Task Force (FATF) (2021): *Opportunities and challenges of new technologies for AML/CFT*, July.

Financial Reporting Council (2024): *Technical actuarial guidance – models*, October.

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

——— (2023): *Enhancing third-party risk management and oversight – a toolkit for financial institutions and financial authorities*, December.

——— (2024): *The financial stability implications of artificial intelligence*, November.

Financial Stability Oversight Council (FSOC) (2023): *Annual Report*, December.

FinRegLab (2021): *The use of machine learning for credit underwriting*, September.

Forrester (2023): *The total economic impact of IBM Watson Assistant*, April.

G7 (2023a): *Hiroshima Process international guiding principles for organizations developing advanced AI system*, October.

——— (2023b): *Hiroshima Process code of conduct for organizations developing advanced AI system*, October.

——— (2023c): *Hiroshima Process international guiding principles for all AI actors*, December.

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

Garcia Ocampo, D, J Taneja, J Yong and J Zhu (2023): "From clicks to claims: emerging trends and risks of big techs' foray into insurance", *FSI Insights on policy implementation*, no 51, August.

Grobelnik, M, K Perset and S Russell (2024): "What is AI? Can you make a clear distinction between AI and non-AI systems?", *OECD.AI*, 6 March.

Gulley, A and A Hilliard (2024): *Lost in transl(A)t(I)on: differing definitions of AI*, February.

Hong Kong Monetary Authority (HKMA) (2022): "Artificial intelligence-based regtech solutions", *Regtech Adoption Practice Guide*, no 6, April.

——— (2024a): *Consumer protection in respect of use of generative artificial intelligence*, August.

——— (2024b): *Generative artificial intelligence in the financial services space*, September.

IIF-EY (2023): *IIF-EY annual survey report on AI/ML use in financial services: public summary*, December.

International Association of Insurance Supervisors (IAIS) (2023a): *Issues paper on insurance sector operational resilience*, May.

——— (2023b): *Regulation and supervision of AI/ML in insurance: a thematic review*, December.

——— (2024a): *Draft application paper on the supervision of artificial intelligence*, November.

——— (2024b): *Holistic framework for systemic risk in the insurance sector global monitoring exercise December 2024* (forthcoming), December.

International Monetary Fund (IMF) (2023): "Generative artificial intelligence in finance: risk considerations", *Fintech Notes*, August.

International Organization for Standardization (ISO) (2023): *Information technology – artificial intelligence – guidance on risk management*, ISO/IEC 23894:2023, February.

JPMorgan Chase (2024): "2024 Investor day transcript", *JPMorgan Chase & Co*, May.

Koh, T Y and J Prenio (2023): "Managing cloud risk – some consideration for the oversight of critical service providers in the financial sector", *FSI Insights on policy implementation*, no 53, November.

KPMG LLP (2023): *The generative AI advantage in financial services*, August.

Ladva, P and A Grasso (2024): "The evolution of AI in the insurance industry", 2 May.

Marsch & McLennan Companies (2019): *Artificial intelligence applications in financial services: asset management, banking and insurance*.

McKinsey (2024): "Scaling gen AI in banking: choosing the best operating model", 20 March.

MIT Sloan Management Review and Boston Consulting Group (MIT-BCG) (2023): "Building robust RAI programs as third-party AI tools proliferate", findings from the 2023 Responsible AI Global Executive Study and Research Project, 20 June.

Monetary Authority of Singapore (MAS) (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 (2023): "Use of artificial intelligence systems by insurers", *NAIC Model Bulletin*, December.

National Institute of Standards and Technology (NIST) (2021): "Four principles of explainable artificial intelligence", *Interagency or Internal Report 8312*, September.

——— (2023): *Artificial Intelligence risk management framework (AI RMF 1.0)*, January.

New York Department of Financial Services (2024a): "Use of artificial intelligence systems and external consumer data and Information sources in insurance underwriting and pricing", *Insurance Circular Letter* no 7, July.

——— (2024b): "Cybersecurity risks arising from artificial intelligence and strategies to combat related risks", *Industry Letter*, October.

NVIDIA (2024): "State of AI in financial services", *Survey Report*.

OECD (2021): *Artificial intelligence, machine learning and big data in finance: opportunities, challenges, and implications for policy makers*, August.

——— (2023): "Generative artificial intelligence in finance", *OECD Artificial Intelligence Papers*, no 9, December.

——— (2024a): "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, no 8, March.

——— (2024b): "Regulatory approaches to artificial intelligence in finance", *OECD Artificial Intelligence Papers*, no 24, September.

Oracle (2024): *Anti–money laundering AI explained*, August.

Perez-Cruz, F and H S Shin (2024): "Testing the cognitive limits of large language models", *BIS Bulletin*, no 83, January.

Prenio, J (2024): "Peering through the hype – assessing suptech tools' transition from experimentation to supervision", *FSI Insights on policy implementation*, no 58, June.

Prenio, J and F Restoy (2022): "Safeguarding operational resilience: the macroprudential perspective", *FSI Briefs*, no 17, August.

Prenio, J and J Yong (2021): "Humans keeping AI in check – emerging regulatory expectations in the financial sector", *FSI Insights on policy implementation*, no 35, August.

Qatar Central Bank (2024): *Artificial Intelligence guideline*, September.

Reinsurance Group of America (RGA) (2024): *Global claims fraud survey*.

Stanford University (2024a): *Artificial intelligence index report 2024*, April.

——— (2024b): *The foundation model transparency index v 1.1*, May.

Statista (2023): "Global generative AI revenue 2023 | Statista", October.

——— (2024): "Financial sector AI spending worldwide 2023, with forecasts to 2027", 19 June.

The Economist Intelligence Unit (2022): "Banking on a game-changer: AI in financial services".

UK Government (2024): "International scientific report on the safety of advanced AI", *Interim Report*, May.

UNESCO (2022): *Recommendation on the ethics of artificial intelligence*, January.

United Nations (2024a): *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*, UN Resolution A/78/L.49, March.

——— (2024b): *Governing AI for humanity*, September.

US Department of the Treasury (2024): *Managing artificial intelligence-specific cybersecurity risks in the financial services sector*, March.

Veritas Initiative (2023): "FEAT Principles assessment case studies", *Veritas Document 6*, June.

West, S (2023): "Competition authorities need to move fast and break up AI", *Financial Times*, 17 April.

World Economic Forum (2024): "Global cybersecurity outlook 2024", *Insight Report in collaboration with Accenture*, January.