OECD publishing

SUPERVISION OF ARTIFICIAL INTELLIGENCE IN FINANCE

CHALLENGES, POLICIES AND PRACTICES

OECD ARTIFICIAL INTELLIGENCE PAPERS
January 2026 **No. 54**

OECD
BETTER POLICIES FOR BETTER LIVES

Disclaimers

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Member countries of the OECD.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Photo credits: © Kjpargeter/Shutterstock

Abstract

While most OECD Member countries consider they have appropriate regulations for the use of AI in finance, challenges may arise in the interpretation and implementation of applicable AI regulations by financial supervisors. This paper analyses current supervisory approaches to the use of AI in finance and challenges in overseeing its adoption. The paper also reviews supervisory practices that balance promoting responsible AI adoption in finance with policy objectives of financial market stability and integrity, and the protection of financial consumers.

This paper is part of the series “OECD Artificial Intelligence Papers”, https://doi.org/10.1787/dee339a8-en

Table of contents

FIGURES

Acknowledgements

This report provides analysis of current supervisory approaches to Artificial Intelligence (AI) in Finance and discusses reported challenges encountered in the interpretation and implementation of applicable AI regulations by financial supervisors in some jurisdictions. The report discusses policies and supervisory practices that balance the promotion of responsible adoption of AI in finance with policy objectives of stability and integrity of financial markets and protection of financial consumers.

The report has been developed by the Capital Markets and Financial Institutions of the OECD Directorate for Financial and Enterprise Affairs. It was drafted by Iota Kaousar Nassr under the supervision of Fatos Koc, Head of the Financial Markets Unit, and Serdar Çelik, Head of Division. Eva Abbott, Liv Gudmundson and Mathilde Le Pichon provided editorial and communication support.

The authors gratefully acknowledge valuable input and feedback provided by the following individuals and organisations: Sara G. Castellanos, Banco de México; Giuseppe Grande, Banca d'Italia; Mikari Kashima, Bank of Japan; Rohan Paris and Paull Randt, U.S. Department of the Treasury; Jasmine Tan, Australian Securities and Investments Commission.

The report was discussed by the OECD Committee on Financial Markets, chaired by Mr Seiichi Shimizu, Assistant Governor, Bank of Japan, on 11 September 2025. The report constitutes part of the horizontal OECD project on Artificial Intelligence.

Executive summary

The transformative potential of artificial intelligence (AI) innovation, catalysed by advancements in generative AI (GenAI) and large language models (LLMs), is poised to significantly reshape the global financial sector. The finance sector, having leveraged machine learning [ML] models for decades, is progressively exploring and deploying GenAI models, while also exploring Agentic AI capabilities.

In OECD economies existing regulatory requirements remain applicable irrespective of the technology used to deliver a financial service or product, given the technology-neutral principle guiding financial regulation. While regulation provides the foundational legal architecture for financial oversight, supervision is a dynamic and ongoing process through which rules and policies are interpreted in practice to ensure compliance and to identify and assess emerging risks to the integrity and stability of the markets.

Financial supervision therefore serves as the practical enforcement mechanism of financial regulation, ensuring that policies translate into effective oversight and resilient financial markets. It is at this level of practical interpretation and implementation of AI policies in finance that challenges may arise, given the intrinsic characteristics of AI innovation, particularly advanced forms of AI.

This paper analyses current supervisory approaches and examines reported challenges encountered in the supervision of AI in finance by some countries. It also discusses possible supervisory practices that balance the promotion of the responsible adoption of AI in finance with policy objectives related to the stability and integrity of financial markets and the protection of financial consumers. The report builds on earlier work of the Committee on Financial Markets on *Regulatory Approaches to AI in Finance* and draws on input from the 2024 OECD Survey on AI in finance and subsequent contributions.

Supervisory approaches to AI in finance

Despite differences in supervisory approaches (such as reliance on legacy frameworks in some regions or the development of AI-specific guidelines and supervisory expectations in others) supervisory efforts are anchored in common principles, notably a risk-based and technology-neutral approach. Where new policy frameworks have been introduced explicitly for AI in finance, layering AI-specific frameworks on top of pre-existing sector-specific rules may complicate the application of supervisory mandate.

Efforts are therefore needed to promote and pursue streamlining and simplification of regulation, as well as clarity and consistency in supervisory interpretation, where necessary. This includes identifying any overlaps, conflicts or inconsistencies in regulations applicable to AI, and clarifying the interpretation of these rules for the purposes of AI supervision, if and where necessary, with a view to assisting supervised entities in their compliance efforts.

Reported challenges in the supervision of AI in finance

Some of the most prominent reported challenges in the supervision of AI in finance relate to distinctive characteristics of AI innovation, such as the pace of its evolution (at least thus far), the opaqueness and complexity of the underlying technology, and novelties related to its dynamic nature and the potentially high degree of autonomy it could, in theory, entail. The lack of comprehensive data on AI adoption by financial services firms complicates the assessment of its use and may pose challenges for monitoring associated vulnerabilities. This is further exacerbated by the growing significance of non-supervised entities, such as third-party technical vendors, many of which operate outside the scope of formal oversight by financial regulators.

Challenges reported by authorities largely mirror the compliance challenges identified by supervised entities, which may in turn impede the wider deployment of AI by financial firms. These challenges include model risk management, validation and compliance assessment of increasingly complex AI systems, as well as limitations related to model explainability. They also include limited transparency and associated considerations around assessing robustness and upholding the fairness of model outputs; alongside governance and data-management challenges. For example, articulating how the concept of 'human in the loop' should apply in practice, depending on the context, is reported as challenging by some authorities.

While existing requirements continue to apply and supervised entities are expected to take AI-specific aspects into account and adapt their risk-management frameworks accordingly, potential guidance and clarification on how compliance requirements align with advanced AI models' technical specificities could be beneficial in some jurisdictions. Depending on the case, such guidance could address any perceived ambiguity as to the way model risk management frameworks should be interpreted and operationalised given, for example, the lack of explainability and the dynamic adaptability and recalibration of AI models. Rather than imposing rigid or overly prescriptive requirements that could inadvertently hinder the adoption of AI innovation, it may be more effective to consider providing interpretative guidance and practical clarifications on the application of existing model risk management frameworks in AI contexts if and when this is deemed necessary.

Balancing policy objectives with the promotion of the responsible adoption of AI

In jurisdictions where supervised entities report challenges arising from a perceived lack of clarity, particularly in light of the overlay of newly adopted AI regulation, carefully calibrated guidance on the interpretation of high-level principles could be beneficial. Additional clarifications could help provide legal certainty for firms, which in turn may strengthen confidence and encourage further investment in responsible AI innovation. Greater clarity for the finance industry regarding regulatory requirements and how to meet them through supervisory expectations and guidance could be beneficial in such jurisdictions. Such guidance could help market participants ensure regulatory compliance and reduce perceived regulatory uncertainty, and support more effective oversight and consistent regulatory outcomes.

Any guidance provided should be very carefully designed and calibrated, to avoid a negative effect on AI adoption by impeding firms' ability to flexibly explore using new technologies. Overly prescriptive approaches should be avoided, given the rapid pace of technological innovation, and a risk-based approach may be most effective for enabling financial institutions to address key risks where needed.

Enhanced forms of proactive engagement between supervisors and industry stakeholders, beyond standard supervisory activities, could foster mutual understanding. Close and sustained engagement with industry can yield significant benefits for supervised entities, while also improving authorities' understanding of the challenges encountered by supervised entities in their compliance efforts. Proactive engagement with the industry through AI-specific testing, such as sandboxes[1] or model testing, can provide the confidence and clarity needed to encourage innovation while protecting markets and their

participants and safeguarding stability. Novel initiatives involving model testing can cultivate productive dialogue between firms developing or deploying AI models and supervisory bodies, fostering mutual understanding and supporting model validation (e.g. the UK Financial Conduct Authority AI Live Testing).

Increased capacity and upskilling of financial supervisors will also be necessary to achieve monitoring and oversight objectives, as well as to enable authorities to develop and deploy AI as part of the supervisory activity *inter alia* through SupTech tools that incorporate AI innovation. Co-ordinated efforts among supervisory authorities could enable the strategic pooling of expertise and institutional capacity in relation to AI-based SupTech tools. Investment is also required to conduct further research into the potential long-term impacts of AI on financial market structures, competition, and financial stability.

Maintaining a flexible, agile and adaptive stance to the financial supervision at the practical level could help address the supervisory challenges discussed above while allowing oversight to keep pace with technological advances. In some cases, allowing for technology-specific guidance or the consideration of novel methodologies and techniques to enrich the supervisory toolkit could assist supervisors in achieving a balance between fostering innovation and ensuring stability. Public-private dialogue between supervisors and regulated entities should accompany the consideration of new supervisory methods, techniques, and tools. Continuous assessment of the supervisory landscape is necessary to ensure that it remains fit for purpose. Supervision should also remain open to enriching and adapting this framework to reflect the realities and specific characteristics of AI innovation, in order to foster responsible AI innovation while mitigating associated risks.

1 Translating policies into effective oversight for AI in finance

1.1. Supervisory approaches to AI in finance

Regulation provides the foundational legal architecture for financial oversight. OECD analysis indicates that most jurisdictions consider they have appropriate regulation in place for the use of AI in finance (OECD, 2024[1]). This includes pre-existing regulation, newly introduced product regulation or cross-sectorial rules, as well as non-binding policy guidance and national policy frameworks which either specifically target financial activities or apply across sectors, including finance. It is also important to note that the vast majority of responding jurisdictions do not plan to introduce new regulations for AI in finance.

Given the technology neutral principle guiding financial regulation, existing rules and guidance remain applicable regardless of the underlying technology used to deliver a financial service or product. This includes laws and regulations on prudent business practices, consumer and investor protection, cybersecurity, and operational resilience, among other areas (see Box 2.1). Many of the risks related to AI are not necessarily new or unique to AI innovation but are instead exacerbated or amplified by the use of such innovation, or manifest in different ways (OECD, 2021[2]; 2023[3]). Advances in technology do not render existing safety and soundness standards or compliance requirements obsolete. This is particularly true in the financial sector, where the use of models has long been integral to the business strategies of market participants, spanning several decades.

Financial supervision operates alongside financial regulation to ensure compliance, stability, and integrity within the financial system and goes beyond the simple enforcement of regulations to include the management of risks. While financial regulation establishes the rules and standards that financial institutions must follow, supervision actively monitors, assesses, and enforces these requirements. Financial supervision therefore serves as the practical enforcement mechanism of financial regulation, ensuring that policies translate into effective oversight and resilient financial markets. It is at this level of practical interpretation of AI policies in finance that supervisory challenges may arise.

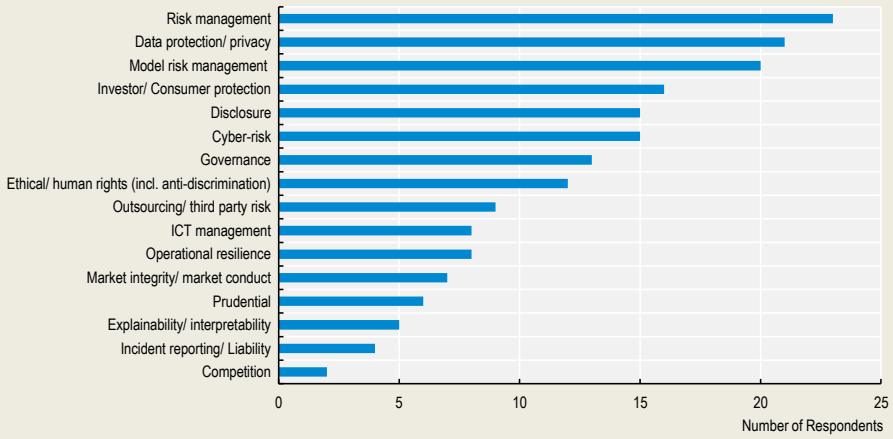
Supervisory approaches to AI in finance vary across jurisdictions, ranging from leveraging existing principles-based frameworks (e.g. UK's framework) to developing AI-specific guidance (e.g. Monetary Authority of Singapore FEAT framework) and integrating cross-sectoral AI regulatory requirements into certain areas of financial supervision (e.g. EU AI Act). Indicatively, in the UK, authorities primarily rely on established principles-based frameworks to guide oversight. Singapore's Monetary Authority has developed dedicated AI governance principles tailored to guide the sector into addressing specific challenges posed by AI through enhanced governance. In the EU, the AI Act incorporates AI-specific requirements for use cases deemed high-risk in insurance and banking, within a broader cross-sectoral regulation, which will need to be incorporated into existing supervisory strategies.

Box 1.1. Regulatory approaches to AI in finance

In September 2024, the Committee on Financial Markets released an overview analysing different regulatory approaches to the use of AI in finance in 49 OECD and non-OECD jurisdictions based on a dedicated Survey on Regulatory Approaches to AI in Finance.

The OECD analysis provides examples of rules and regulations that may apply to the use of AI in finance and that can be grouped under a set of areas as depicted in Figure 1.1. Most of the covered areas relate to risk management and, in particular, model risk management; data-related frameworks; consumer and investor protection, as well as governance and accountability requirements.

Figure 1.1. Examples of areas covered by existing financial sector rules



Note: Non-exhaustive, as reported by respondents to the survey.
Source: OECD (2024[1]), Regulatory approaches to Artificial Intelligence in finance, https://www.oecd.org/en/publications/regulatory-approaches-to-artificial-intelligence-in-finance_f1498c02-en.html

Across jurisdictions, different forms of binding and/or non-binding policy instruments have emerged to complement existing financial regulations in response to AI advances. Some countries have enacted cross-sectoral legislation encompassing financial activities (e.g. EU AI Act and national laws in Brazil, Colombia, and Peru), while others have pursued targeted regulatory proposals focused on specific actors and activities. In parallel, about c. a quarter of respondents to the OECD survey have issued non-binding guidance, including blueprints, principles, and white papers, either at the cross-sectoral level or tailored to financial domains. These instruments generally aim to establish priorities and promote safe and responsible AI innovation. Despite variations in format, there are significant commonalities in content, emphasising fairness, accountability, ethical use, compliance, transparency, and robust governance mechanisms. Some jurisdictions additionally urge financial authorities to leverage their full supervisory remit to mitigate AI-related risks to consumers and investors. Importantly, the different approaches outlined above are not mutually exclusive.

Source: OECD (2024[1]), Regulatory approaches to Artificial Intelligence in Finance, https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/09/regulatory-approaches-to-artificial-intelligence-in-finance_43d082c3/f1498c02-en.pdf

Although supervisory approaches to AI in finance may appear disparate across jurisdictions, they are all underpinned by the same foundational principles, in particular the risk-based approach to supervision and a technology neutral/agnostic stance to innovation. Under risk-based supervision, supervisory resources and interventions are prioritised according to the relative risk profile of financial institutions or sectors, with the intensity of supervision aligned to the prevailing risks to which these are exposed. Rather than applying uniform oversight, risk-based supervision enables supervisors to focus more intensively on entities or activities that pose higher risks to financial stability, consumer protection, or market integrity. In addition to enhancing the effectiveness of supervisory efforts by tailoring their monitoring, inspection, and enforcement strategies accordingly, risk-based supervision can also support financial inclusion by reducing regulatory burdens on lower-risk entities or activities.

Technology neutrality, which serves as a foundational element of financial regulation in OECD countries, is also reflected in oversight practices. Financial supervisors aim to apply consistent oversight standards irrespective of the technologies employed, thereby ensuring that regulation remains resilient and adaptable in the face of innovation. When AI is used in areas covered by existing rules or guidance, such rules or guidance should generally apply, whether decisions are made by AI (with or without human intervention), traditional models, or humans (OECD, 2024[1]). This transposition enables regulatory frameworks to foster competition and innovation without prescribing or favoring specific technological solutions, while maintaining their core objectives of market integrity, consumer protection, and financial stability.

Nevertheless, at the level of practical implementation, financial supervisors have reported tangible challenges in effectively translating and interpretating technology-neutral regulatory provisions across specific domains associated with the use of AI in finance (Section 4.1). These challenges frequently stem from the distinctive characteristics and novel dimensions of AI innovation, particularly its increasing complexity and rapid pace of evolution. Supervisory challenges are also reported as resulting from the interplay between sectorial rules and new regulation or specific AI-related guidance, where these have been established. Additional challenges relate to evolving institutional and market structures, for example, due to the growing role of technology providers in the deployment and scaling of AI by financial market participants. Additionally, the limited availability of granular data on the current state of AI adoption can further complicate monitoring activities and may hinder the effective management of identified risks.

1.2. Oversight frameworks: interplay between sectorial rules and other policies

Supervisory challenges may arise from the interplay between existing tech-neutral sectoral regulations and emerging cross-sectorial or finance-specific AI policy frameworks, where such rules have been formulated. Depending on the jurisdiction, oversight of financial activities involving the use of AI could require the interpretation of a combination of pre-existing financial sector regulation, cross-sectorial policies and practices (e.g. anti-discrimination practices and ethics-related rules such as the US Interagency Fair Lending Examination Procedures (FFIEC, 2009[4]), product-safety or other cross-sectorial regulation introduced explicitly for AI with applicable financial use cases (e.g. EU AI Act (EU, 2024[5])), and newly introduced sectorial rules, guidance or principles for (parts of) the financial sector (e.g. guidance on data ethics within the insurance sector in Ireland (Central Bank of Ireland, 2023[6]).

For example, in the EU, the use of AI in the financial sector beyond the cases identified as high-risk[2] and any other use cases included in the EU AI Act (e.g. onboarding), will be addressed in accordance with pre-existing legislation) applicable irrespective of technology used, including the Markets in Financial Instruments Directive (MiFID) II or Solvency II. By way of an example, MiFID II includes requirements for investments firms and trading venues engaged in algorithmic trading and high-frequency trading (HFT), activities that can also be based on AI systems (ESMA, 2024[7]). Similarly, the use of third-party AI systems could be considered outsourcing under Solvency II and insurance undertakings in the EU remain fully responsible for all their obligations under Solvency II for outsourcing (EIOPA, 2024[8]). Where new policy

frameworks have been introduced explicitly for AI in finance, layering AI-specific regulatory frameworks on top of pre-existing sector-specific rules may complicate the application of supervisory mandate and/or increase the complexity of financial firms' compliance efforts, introducing potential perceived ambiguity. Efforts are therefore needed to promote streamlining, clarity and consistency in supervisory interpretation. This includes identifying any overlaps or inconsistencies, as well as clarifying how these rules should be interpreted for the purposes of AI supervision, in order to assist supervised entities in their compliance efforts. Where new rules have been introduced, authorities could also consider integrating requirements targeting the use of AI into existing sectorial frameworks, with a view to simplify both the oversight activity and the compliance efforts of supervised entities. For instance, in responding to the Australian Government's preliminary consultation on implementing high-risk AI guardrails, the Australian Securities & Investments Commission (ASIC) highlighted the importance of reconciling and aligning new legal guardrails with overlapping laws, particularly where new requirements could impose lower standards of conduct on regulated entities or where the allocation of legal responsibilities along the AI supply chain could conflict with existing legal obligations (ASIC, 2024[9]).

These efforts are also important in the context of ongoing work by financial regulatory authorities to review existing policy frameworks to ensure they remain fit for purpose. Indeed, a majority of OECD countries indicated that there may be a need to evaluate whether any further strengthening or expanding existing rules beyond those already in place would be necessary, or useful, to achieve regulatory authorities' policy objectives. This includes the need to assess the possible interplay between existing and new rules that may require adjustment or expansion following their implementation. This could also be beneficial if, for example, any gaps in risk mitigation are identified over time under existing rules, or if there is a need tore-interpret existing rules or provide additional guidance to better support the delivery of policy objectives. All the above should be undertaken with a forward-looking approach, taking into account future developments in AI innovation, which is expected to continue transforming financial services at increasing scale and complexity.

1.2.1. Supervisory architecture for AI in finance and supervisory coordination

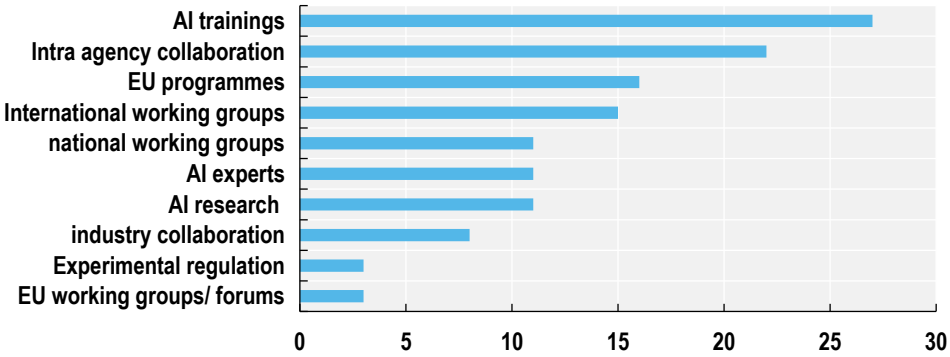
A possible complexity is linked to the evolving supervisory architecture in certain jurisdictions, where the institutional arrangements for the oversight of AI-related innovation are expanding at the government or supervisory authorities' level to encompass additional layers of direct or indirect oversight of financial activities. In Europe, in particular, the designation of market surveillance authorities (MSA) for AI, although likely to include financial supervisory bodies as designated MSAs, adds a new layer of oversight to existing mandates, especially for enforcing the EU AI Act in relation to high-risk use cases and prohibited AI practices directly linked to financial services or products. At the regional level, the AI Office, the AI Board and the ESAs are also involved in AI governance, alongside national authorities. In other jurisdictions without newly introduced AI monitoring bodies, other digital authorities could have cross-sectoral responsibilities for the oversight of AI activity, including in finance, depending on the jurisdiction. For instance, in Singapore, the Infocomm Media Development Authority (IMDA), plays an important role in shaping and overseeing AI governance across sectors, including in finance, alongside the Monetary Authority of Singapore (MAS)[3]. In Australia, ASIC identified some of the specific challenges associated with the introduction of a new AI regulator (or existing regulator with expanded remit). These include increased complexity for: (a) existing regulators when enforcing laws where there are overlapping remits or unclear boundaries, or where the regulated population of a new regulator is not clearly defined; (b) organisations, in understanding how multiple regimes apply; and (c) consumers, inn determining how to seek redress when harms occur.

Recognising the cross-cutting nature of AI, which increasingly involves authorities beyond the financial sector, financial authorities have highlighted that coordination and information sharing are vital for the safe adoption, development and deployment of AI. Authorities have reported various forms of supervisory coordination efforts (Figure 1.2), involving coordination and information-sharing working groups at national

and international levels between financial authorities; inter-agency working groups; engagement with AI expertise in the private sector and with academia; but also, supervisory training initiatives.

The examples of supervisory coordination efforts reported by authorities underscore the critical importance of institutional collaboration. These coordinated efforts aim at building a collective understanding of emerging risks and best practices to mitigate these, coordinating possible enforcement action particularly when it comes to cross-border activities, as well as facilitating more coherent oversight frameworks in response to the rapidly evolving AI landscape.

Figure 1.2. Supervisory coordination efforts



Note: Non-exhaustive, as reported by respondents to the survey.
Source: OECD 2024 Survey on Regulatory Approaches to AI in Finance.

Supervisory collaboration and coordination will become increasingly important for effective oversight as AI diffusion intensifies in the financial sector (see Section 3.4). Joint efforts can allow authorities to share insights, methodologies, and best practices, reducing duplication and enhancing collective capacity. Collaboration at the national level can support inter-agency efforts for simplification, address overlaps and ensure consistency in the application of policy frameworks applicable to AI. Cross-border cooperation can support consistency and provide legal certainty in the deployment of AI systems that traverse national barriers, while mitigating risks of regulatory arbitrage. At the operational level, coordinated efforts among supervisors can enable the pooling of institutional capacity and expertise, as the supervision of complex AI systems increasingly relies on technical expertise in addition to traditional financial sector expertise (see Section 3.4).

1.3. Data gaps and monitoring tools

Effective risk monitoring requires timely and granular information on the use of AI systems in finance, as well as metrics related to identified sources of risks associated to such use. However, financial authorities are still at an early stage in their efforts to systematically monitor AI adoption within the sector. Regulators and supervisors have taken different approaches to data collection on AI deployment, through dedicated market surveys, the incorporation of AI-specific questions in sectorial surveys, and in the context of their ongoing monitoring activities.

The lack of comprehensive data on AI adoption by financial services firms complicates the assessment of AI usage and poses challenges for monitoring vulnerabilities and potential financial stability implications (FSB, 2024[10]). Despite progress in data collection, important challenges persist, including a lack of standardised definitions and taxonomies, fragmented information sources, high data collection costs, and difficulties assessing data relevance, materiality, and criticality to specific AI use cases. The growing

complexity and integration of AI systems within processes and workflows may also hinder effective mapping and monitoring. These issues could further limit oversight effectiveness and challenge the development of evidence- and risk-based supervisory approaches. In the context of financial stability, the FSB has identified a range of direct and proxy indicators that can support the monitoring of AI adoption and related vulnerabilities in the financial system (FSB, 2025[11]). Similar efforts may be desirable in other areas of oversight and supervision, and coordinated data collection efforts could help standardize information gathering and formalize common metrics.

At the international level, common understanding of definitional differences (e.g. General Purpose AI) can assist in international comparability of emerging trends and strengthen cross-border supervisory coherence. This is increasingly important given the global nature of AI innovation and its diffusion across interconnected financial markets. From a financial stability standpoint, the FSB has identified a range of direct and proxy indicators that can be helpful in supporting the monitoring of AI adoption and related vulnerabilities in the financial system (FSB, 2025[11]).

The lack of a common supervisory language for AI in finance warrants careful attention from policymakers. Given the inherently multidisciplinary nature of AI oversight, effective cooperation among stakeholders, including legal experts, economists, statisticians, scientists and computer engineers, depends on shared terminology and a common conceptual understanding. Closely related challenges arise in data collection, notably the absence of standardised definitions and taxonomies, fragmented sources, high costs, and difficulties in assessing data relevance. Disseminating clear definitions and taxonomies for key AI concepts through guidance and interpretative clarifications may help address these challenges.

1.3.1. Increasing importance of non-supervised entities in financial activity

The increasing reliance of financial market participants on third-party providers for AI-related services (e.g. AI models and cloud infrastructure) raises the risk of third-party dependency on a concentrated number of vendors (OECD, 2024[1]; FSB, 2024[10]). At the same time, it should be acknowledged that the efficiencies and scalability of AI innovation in finance are built upon, thanks to the services provided by such-third party providers.

The growing significance of non-supervised entities, including third-party technical vendors, within financial markets, introduces risks for supervised entities. Reported risks related to the increasing reliance of the financial sector on these third-party providers include deficiencies in transparency and control, potential vendor lock-in, concentration risk, and operational risks related to model maintenance or business continuity.

Depending on the jurisdiction, financial regulatory authorities may not have any supervisory powers over these third-party providers that fall outside the regulatory perimeter. This can result in unassessed risks or difficulties in monitoring emerging vulnerabilities, to the extent that such risks are not mitigated through supervised entities' risk management of third-party relationships. In some cases, such as those in EU countries, financial supervisors are mandated to identify and designate critical ICT third-party providers for financial services. These entities are then subject to direct EU-level oversight, including audits, inspections, and compliance reviews. Additionally, in the EU, General purpose AI providers will be supervised by the European Authorities (AI Office) for high-risk use cases identified under the EU AI Act.

Regulators are likely to encounter challenges arising from disparate supervisory approaches, particularly when different types of regulators supervise different parts and actors in the AI supply chain.

In some jurisdictions, supervisors are managing the growing relevance of third-party risk management by increasingly shifting the focus on the supervised entities' capabilities in managing these dependencies effectively. This includes the ability of supervised entities to conduct thorough due diligence, impose adequate contractual controls (including liability provisions) and perform meaningful testing on external

systems. Contingency planning for vendor failure or unavailability can also serve as a proxy for ensuring the safety and soundness of externally sourced AI capabilities in this context.

Supervisors generally expect supervised entities to manage the risks associated with third-party relationships, and regulators may supervise these relationships. Instruments such as the FSB toolkit for third-party risk management and oversight can help financial institutions identify critical third-party services and manage risks throughout the lifecycle of a third-party service relationship, while guiding supervisors in their monitoring activities (FSB, 2023[12]). In addition to evaluating supervisory and audit data related to outsourcing of ICT services, financial supervisors in some jurisdictions are also performing specific checks and on-site reviews at both supervised financial institutions and service providers (FINMA, 2024[13]).

Similar concerns may arise in relation to the adoption of AI-driven tools by certain types of non-bank financial institutions (NBFIs), given their increasing role in the global financial system; associated data gaps or risks of regulatory arbitrage. AI could also increase the relative footprint of these institutions in certain markets, for example through expanded use of algorithmic trading (Bank of England, 2025[14]).

2 Potential challenges to supervision of AI in finance

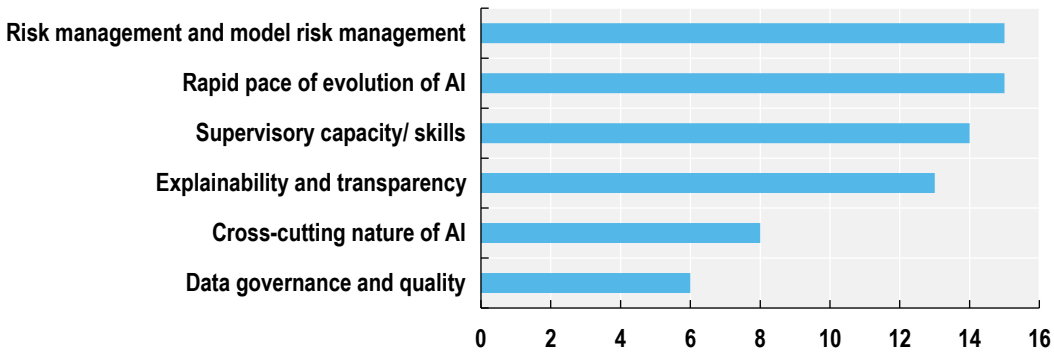
2.1. Reported challenges to AI supervision in finance

Supervisory authorities must grapple with AI systems that are increasingly integrated into the processes of financial institutions, are dynamic in nature and are often opaque in their functioning, potentially posing challenges for traditional oversight mechanisms and the interpretation of regulatory frameworks. Specific challenges have been reported in areas such as risk management and model risk management frameworks; explainability and transparency of AI-driven models; data management frameworks; as well as supervisory capacity (see Section 4.3). Intrinsic characteristics of AI innovation, such as its complexity, rapid pace of development, and cross-cutting nature, are additional challenges reported by financial authorities.

The reported challenges are particularly pronounced in areas of AI that are subject to supervisory expectations. In that sense, supervisory challenges largely mirror compliance challenges, and vice-versa, compliance difficulties may also lead to specific challenges for supervisory authorities. These relate primarily to the opaqueness of advanced AI models, the ensuing explainability and transparency issues, coupled with related governance considerations, particularly when discussing reliance on third-party models and infrastructure, as well as to capacity and skills.

Understanding these and other challenges faced by supervisors is critical for several reasons. Effective oversight is paramount for maintaining financial stability, ensuring market integrity, and protecting consumers and investors from harm, such as bias, fraud, or data misuse, while fostering an environment in which responsible AI innovation can thrive without compromising core policy objectives. Identifying and analysing these challenges can inform policy development, guide resource allocation within supervisory agencies, and highlight areas requiring greater international cooperation and standard-setting.

Figure 2.1. Identified challenges in supervision of AI in finance



Note: In number of jurisdictions. Non-exhaustive, as reported by respondents to the survey.
Source: OECD 2024 Survey on Regulatory Approaches to AI in Finance.

2.1.1. Distinctive characteristics of AI innovation: speed, complexity, autonomy

The unique nature of AI innovation, characterised by advanced capabilities and rapid pace of progression, is fuelling its potential impact on economic growth and productivity. These distinctive characteristics and transformative effects position AI alongside earlier General Purpose Technologies (GPTs), such as the internet and personal computers, as well as historic breakthrough innovations such as the steam engine and electricity (Filippucci et al., 2024[15]; Filippucci, Gal and Schief, 2024[16]).

The flipside of this for supervisory activity is also increasingly recognised, particularly the challenges of keeping pace with the rapid speed of AI advances and maintaining effective oversight in the face of this rapidly evolving technology. AI's velocity and its impact on the sector often outpaces supervisory efforts, potentially risking that oversight lags behind or even becomes ineffective in certain cases. Supervisors have to interpret principles-based rules, ensuring that the oversight framework remains relevant and effective in a dynamic and increasingly complex environment.

The potential for a higher degree of autonomy in the future, as in the case of Agentic AI[4], serves as a further catalyst for efficiencies and productivity gains, but would also raise significant challenges in terms of governance, accountability and monitoring for both practitioners and supervisors. These characteristics are particularly pertinent to financial oversight, given its inherent forward-looking nature: financial supervisors aim to identify and mitigate risks associated with, or amplified by the use of AI, and to intervene proactively to address potential threats.

The existence of many 'unknown unknowns' (e.g. unanticipated vulnerabilities and hidden linkages) and the difficulties in understanding how AI will react to such uncertainty adds an extra layer of complexity, which is also associated with trust in AI usage (Daníelsson, Macrae and Uthemann, 2022[17]). While supervisors cannot predict such 'unknown unknowns', they are well-equipped to respond to them with well-established processes and practices, assisted by their experience in managing crises. That said, agility and ability to adjust and adapt supervisory practices and methodologies will also be important in light of AI's distinctive nature. According to some regulators, the pervasive, cross-cutting nature of AI innovation may call for robust macro surveillance of interdependencies and strong cross-border supervisory collaboration to allow for the benefits of responsible AI to materialise across sectors and economies in a safe manner.

2.2. Model risk management, validation and compliance assessment

The use of models in the financial sector predates the advent of advanced AI, and comprehensive frameworks, standards and guidance for model governance, transparency, governance and accountability, and operational resilience have been in place for the last two decades. Model risk management is indeed the prime example of an area where existing regulations, guidance and general frameworks for risk and model risk management continue to apply to AI-based models (see Box 3.1). These encompass validation protocols, performance monitoring practices and regulatory compliance mechanisms that are imperative for the use of any type of model in finance (and beyond). These frameworks are designed to be sufficient to identify, manage, monitor and control the risks associated with both simple and complex models and are applicable to AI models in OECD economies with technology-agnostic financial regulation. However, some jurisdictions report some confusion among supervised entities about whether model risk management frameworks may apply to all use cases of AI. Supervised entities' use of AI can range from comprehensive AI models to the occasional use of AI as a "tool" to supplement existing work. There is some debate among supervised entities as to what extent the governance frameworks for models may apply.

Box 2.1. Examples of existing model risk management frameworks applicable to AI-driven models

In the EU, investment firms and exchanges engaging in algo-trading are required under MiFID II regulation, to put in place effective systems and risk controls, in order to prevent errors and ensure resilience and proper functioning of their trading systems. Under the same regulation, trading venues are required to have systems to ensure resilience, stability and reliability (capacity to deal with peak order volumes, test to ensure continuity of services, install circuit breakers and carryout stress tests) (EU, 2014[18]). Also, requirements on IT and management of ICT risks are foreseen by the new EU harmonised regulatory framework on digital operational resilience anchored in the new Digital Operational Resilience Act (DORA) (EU, 2022[19]).

In the UK, the Bank of England Prudential Regulation Authority's Supervisory Statement 1/23 'Model Risk Management principles for Banks' and its principles on model identification and model risk classification provide a comprehensive framework (Bank of England, 2023[20]). Similar principles are issued by the German financial regulators covering model choice, data, validation, and explainability, underscoring that the nature of the requirements is intentionally high-level as it is intended to be applicable to all model types used in Finance, including AI/ML models (Deutsche Bundesbank and BaFin, 2021[21]). BaFin's Minimum Requirements for Risk Management (MaRisk) framework provides proportionate, technology-neutral rules, specifying that implementation should reflect the complexity of the model, the risks associated with its use, and its relevance within overall risk management (BaFin, 2023[22]). General IT requirements are also applied on a risk-based and proportionate manner. Without explicitly referring to AI, given the technology-neutral principle, these requirements are expected to cover AI use cases through proportionate assessment of model complexity.

In the US, risk-management principles articulated by the federal banking regulators – the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), and the Federal Reserve Board (FRB) – provide a framework for banks using AI to operate in a safe, sound, and fair manner, commensurate with the materiality and complexity of the risks associated with the relevant activities or business processes.

Federal financial regulatory agencies have issued various guidance and supervisory materials to help banks apply sound risk management principles, including Supervisory Guidance on Model Risk Management (Federal Reserve, 2011[23]; OCC, 2011[24]; FDIC, 2017[25]). In Switzerland, FINMA has outlined its supervisory expectations regarding the management of key risks, including those related to models (FINMA, 2023[26]).

Of particular importance to some countries are rules for internal rating models used by banks. Where AI is incorporated into such models, they are expected to meet established statistical model validation methods (e.g., in-sample vs out-of-sample tests, bias-free, basic explainability, etc.) based on currently existing rules. In the EU for example, the European Banking Authority has issued guidance on the use of ML in internal ratings-based (IRB) models, discussing also the interaction between prudential requirements on IRB models with the EU AI Act and the GDPR regulation (EBA, 2023[27]), followed by ECB's guide to internal models (ECB, 2024[28]).

In Canada, the Office of the Superintendent of Financial Institutions (OSFI) has issued guidance on different aspects applicable to the use of AI by federally regulated financial institutions (FRFIs) including around model risk management (OSFI, 2024[29]). A proposed updated guideline is currently under consultation and sets out OSFI's expectations related to enterprise-wide model risk management (MRM) built on strong model lifecycle principles. It will apply to all federally regulated financial institutions and to all models, including AI/ML, whether or not they require formal regulatory approval.

Source: Supervisory authorities, OECD 2024 Survey on Regulatory Approaches to AI in Finance.

Nevertheless, the increased complexity and intrinsic opacity of advanced AI models, the absence of explainability, and the limited transparency of some AI systems may present challenges for monitoring and assessing compliance with existing comprehensive model risk management frameworks. The so-called "black box" nature of advanced models and the difficulty in understanding why and how the model produces outputs may complicate supervisory efforts for model validation. The inability to explain how some outcomes are generated, or to deconstruct the internal rationale that guides some model outputs, challenges requirements related to the justification of decision-making (e.g. credit decisions). Such opacity may also complicate auditability and transparency requirements for models which in turn complicates accountability assignment required under most supervisory frameworks. In addition to challenges in terms of assessment of compliance with frameworks and rules, such levels of complexity and opacity may also complicate supervisors' efforts to evaluate systemic risk exposure.

While existing requirements continue to apply and supervised entities are expected to take AI-specific aspects into account and adapt their risk management frameworks accordingly, additional guidance and clarification as to how compliance requirements align with advanced AI models' technical specificities could be beneficial in some jurisdictions. This may be relevant in jurisdictions where supervised entities report facing challenges arising from a perceived lack of clarity, particularly in light of the overlaying of newly introduced AI regulation with pre-existing sectorial regulation. In such cases, carefully calibrated guidance as to the interpretation of high-level principles could be beneficial in providing legal certainty for firms, which in turn can strengthen confidence and encourage further investment in responsible AI innovation.

Depending on the case, such guidance could address any ambiguity as to the way model risk management frameworks could be interpreted or operationalised given, for example, the lack of explainability, the dynamic adaptability and recalibration of AI models[5], its probabilistic nature and output robustness issues (e.g. 'hallucinations'[6] and anthropomorphism[7]), and ethical considerations (e.g. around fairness). Rather than imposing rigid or overly prescriptive requirements that could inadvertently hinder the adoption of AI innovation, it may be more effective to provide interpretative guidance and practical clarifications on the application of existing model risk management frameworks in AI contexts. Such guidance could support more effective oversight in such jurisdictions, while also assisting the industry in its compliance efforts by clarifying how supervisory expectations align with the technical realities of AI, allowing them to calibrate their AI governance frameworks accordingly. Any guidance provided should be carefully calibrated to avoid negatively affecting AI adoption by impeding firms' ability to flexibly adopt new technologies.

Banque de France ACPR's guidance on "Governance of Artificial Intelligence in Finance" provides recommendations on AI design methodology, as well as on the evaluation of model robustness and algorithmic performance (ACPR, 2020[30]). ACPR recommends that the AI engineering process cover the entire model lifecycle and build in reproducibility, quality assurance, monitoring and full traceability of the process to ensure auditability. In terms of assessment of model performance, performance metrics of algorithms should be carefully selected to evaluate technical efficacy of the algorithm and/or its business objectives, taking into account the inherent trade-off between the algorithm's simplicity and its efficacy (ACPR, 2020[30]).

In Japan, the Financial Services Agency advises supervised entities to take into account model risk in the aggregate so as to account for possible interdependence of different models in order to appropriately address risks (FSA Japan, 2021[31]). In Japan, the JFSA published an AI Discussion Paper (DP Version 1.0) in March 2025, outlining supervisory expectations and future directions to support sound AI utilisation by financial institutions, based on surveys and international developments (FSA Japan, 2025[32]). In Korea, the Guideline on the Use of Artificial Intelligence in Financial Services specifies actions to be taken by financial and finance-related institutions offering financial transactions and other customer-oriented services using AI technologies to manage risk through the AI model lifecycle, including planning, design, development and operation phases (FSC Korea, 2021[33]).

In the UK, based on the SS1/23 Model Risk Management principles issued in 2023, banks are expected to conduct model development testing for material model changes, including material changes over a period of time in dynamic models (Bank of England, 2023[20]). In the US, the National Institute of Standards and Technology (NIST) has produced an AI risk management framework (AI RMF), which includes a range of profiles designed to reflect diverse combinations of use cases and sectoral contexts (NIST, 2023[34]).

More recently, the Monetary Authority of Singapore (MAS) released a paper setting out good practices for AI and GenAI model risk management, observed during the review of the country's bank sector (MAS, 2024[35]). The paper focuses on governance and oversight, key risk management systems and processes (AI identification, inventorisation and risk materiality assessment), and development and deployment of AI (including validation, monitoring and change management). MAS encourages all financial institutions to reference these good practices when developing and deploying AI models, emphasizing the need for robust cross-functional oversight, continuously updated policies and procedures, risk materiality assessments based on impact and complexity, rigorous development/validation/ongoing monitoring processes (including data quality, explainability, bias mitigation and independent audits/reviews). The guidance also includes specific controls for GenAI (e.g., limiting their scope, ensuring human oversight), and for diligent management of third-party AI risks through compensatory testing, contingency planning, and strong contractual clauses in legal agreements as well as investment in training and awareness. This guidance builds upon the FEAT (Fairness, Ethics, Accountability, Transparency) principles issued in 2018, further aligning regulatory expectations with the evolving technical realities of AI systems (MAS, 2018[36]).

Robust model validation is of paramount importance in cases where models influence core business functions or have a direct bearing on consumer outcomes. Ensuring the reliability, accuracy, and integrity of such models is also essential to safeguarding operational resilience and maintaining public trust. Several testing and validation methodologies are being explored and implemented by financial market participants. Some of these include pre-deployment checks and independent validation, robustness testing (assessing model performance under different conditions or with noisy data), adversarial testing (probing model weaknesses using intentionally challenging inputs), stress testing tailored for AI vulnerabilities, and continuous monitoring of model performance post-deployment to detect drift or degradation. Implementing and monitoring the outcomes of these methods requires specialised expertise both within financial firms and supervisory bodies. The effectiveness of model validation techniques also depends on access to model details and underlying data, which can be challenging in case there is limited transparency in models developed by, and potentially deployed in collaboration with, third-party service providers. It should also be noted that, for robust governance, it is crucial to retain and consider each model instance, not just the latest version. This approach ensures that if a model exhibits problematic behaviour (for example, bias or discriminatory outputs), and is subsequently updated or fixed, a record remains available for future interrogation. In AI implementation, the absence of such traceability has proven problematic, as issues can arise post-fix and cannot be adequately investigated if only the latest model is available.

2.3. Explainability, transparency and fairness

Limited explainability has been identified by some supervisory authorities as one of the potential challenges at the oversight level of AI in finance (Figure 3.1). The 'black box' nature of some AI models, which is massively exacerbated in advanced GenAI models, raises risks connected to the lack of explainability and interpretability[8], that is, the difficulty in understanding why and how AI-based models generate results (OECD, 2021[2]; 2023[3]). Limited explainability makes it more complex to detect and correct flaws and evaluate robustness of outputs, and could undermine users' trust, by obscuring potential biases and hindering the assessment of fairness in model outputs. Depending on the level of opaqueness, it may be difficult for supervised entities to comply with their regulatory[9] and reporting obligations and/or to explain decision-making processes to supervisory authorities and to customers. In advanced GenAI applications, it may be challenging to achieve traditional auditability by tracing back the code to decisions that are being

taken for audit trails. At the same time, given the possible trade-off between explainability and performance of the model, financial firms are trying to strike the right balance of sufficient transparency to support accountability, trust, and regulatory compliance, while maintaining high levels of model performance, depending on the specific use case/ area of application concerned.

Depending on the circumstances, supervisory guidance could be beneficial in clarifying the operational interpretation of explainability requirements, thereby facilitating their consistent and effective implementation across regulated entities. Such guidance could support supervised institutions in aligning their interpretation and implementation of explainability requirements, thereby mitigating the risk of regulatory divergence and preserving consistency across supervisory authorities (and jurisdictions) particularly in jurisdictions or regions with AI-specific regulation introduced on top of applicable sectorial rules. For example, the EIOPA AI Governance Principles provide guidance on how EU insurance firms could adapt the types of explanations to specific AI use cases and to the recipient stakeholders, in a proportionate risk-based manner (e.g. in particular in high-impact AI use cases). The guidance clarifies that in certain cases insurers may combine model explainability with other governance measures insofar as they ensure the accountability of firms, including enabling access to adequate redress mechanisms, and outlined tools in transparency and explainability to help identifying areas where fairness is threatened (EIOPA, 2021[37]). Similarly, Korea has issued a cross-sectorial 'Explainable AI Guide' in April 2024. In Canada, the proposed updated guideline E-23 requires explanations to be provided in terms of outputs of AI systems, and a process of understanding and communicating the model outcomes (OSFI, 2024[38]).

Given the principles-based nature of the regulatory framework, firms are expected to implement adequate measures to ensure a level of explainability that is commensurate with the materiality of AI applications. For example, the US NAIC AI Bulletin informs insurers that the controls and processes they should adopt and implement as part of their AI Systems Governance Program should be reflective of, and commensurate with, the insurer's own assessment of the degree and nature of risk posed to consumers by the AI systems that they use, considering the transparency and explainability of outcomes to the impacted customer (NAIC, 2023[39]).

In some jurisdictions, supervisors have also clarified that explainability requirements needs to be put into the right contextual framework to define its purpose and guide its assessment. In France, for example, the notion of explainability relates either to end users (whether financial consumers or internal users); or to those tasked with compliance or governance of the model who need to ensure the consistence of workflows where humans make decisions or facilitate validation and monitoring of AI models (ACPR, 2020[30]). The discussion also introduces four levels of explanation (observation, justification, approximation and replication) to clarify the expectations in terms of explainability of AI in finance, depending on the targeted audience and the associated business risk. Such context dictates the level and form of an appropriate explanation for the AI model, which should also be in conformity with the associated governance framework (ACPR, 2020[30]).

The explainability question is also closely associated with possible concerns around transparency, which, in turn, could be exacerbated in case of third-party provision of AI models. The difficulty in understanding the inner workings of AI models could be amplified in third-party vendor-provided proprietary models. This could be due to intellectual property concerns of third-party service providers, to contractual limitations, or to technical constraints, all of which involve lack of visibility and can hinder the assessment of associated risks by firms using these models and by supervisors overseeing activity enabled by them.

Lack of explainability also links to ethical considerations of AI-driven models in finance, such as fairness, by hindering the detection of bias and discrimination. These ethical dimensions are increasingly embedded within supervisory frameworks and are incorporated in supervisory guidance, where this is made available. Indicatively, MAS's FEAT principles explicitly incorporate Fairness, Ethics, Accountability, and Transparency as core expectations. The ECB's approach links AI governance to the fundamental right to good administration, encompassing fairness and transparency. The US Fed and other US agencies have

long-standing policies around fairness and non-discrimination in lending, which extend to AI-driven models (e.g. Fair Credit Reporting Act, Equal Credit Opportunity Act). In general, supervisors assess these ethical dimensions by reviewing firms' fairness testing procedures and results, examining documentation related to model explainability, evaluating the design of governance structures to ensure clear accountability, and checking compliance with overarching laws and regulations (e.g. anti-discrimination).

Nevertheless, verifying the fairness of AI systems and the efficacy of firms' bias detection and mitigation strategies could be a challenge for some supervisory authorities. This is due to the increasing technical complexity and opacity of both the models and the mitigation techniques used by supervised entities to make their self-assessments. To that end, some supervisory authorities are building on their existing supervisory toolkit to incorporate new tools and methodologies that reflect the technical complexities of AI systems.

For example, MAS' Veritas Initiative was launched as a complement to the FEAT principles issued by the authority in 2018 and aims to operationalise those ethical principles for the use of AI and data analytics. The initiative provides a structured framework and practical toolkits to help financial institutions assess and demonstrate alignment with these principles. These include sector-specific guidance and open-source tools that support firms' evaluation of AI models for fairness, transparency, and accountability across various use cases, including credit risk scoring, customer marketing, and fraud detection.

2.4. Governance and data management

Governance, including data management, is identified as another possible challenge for the monitoring of AI activity in finance, but also for firms' compliance efforts with applicable rules. This involves considerations around the expected human role and interaction with the AI system, including articulating how the concept of 'human in the loop' should apply in practice, depending on the context. In line with the risk-based approach to financial supervision, this could also depend on level of criticality of the use case or process. At the same time, it remains essential to highlight the existence of principles such as accountability assigned to human oversight, which apply irrespective of the technological context.

Human involvement is multidimensional and may require reflection on the impact of 'automation bias', observed where humans place excessive trust in the results produced by machines (ACPR Banque de France, 2025[40]). Based on recent experimentation by ACPR BdF, the conversational explanations generated by GenAI-based robo-advisors inadvertently heightened users' confidence in the guidance provided by the tool, even when the advice was incorrect. Interestingly, explanations did not significantly improve user's understanding or their ability to follow the advice, depending on whether it was correct or not (ACPR Banque de France, 2025[40]). Such results may also be pertinent when it comes to the use of AI by supervisory authorities in the execution of their oversight duties, for example as assessment tools or support for decision-making.

Establishing clear lines of responsibility for the decisions and actions taken by AI systems could become challenging as parts of the AI systems become more autonomous (e.g. Agentic AI). However, requirements for governance frameworks ensure that accountability is effectively assigned and managed. Any perceived ambiguity in terms of the accountable party (developer, deployer, data provider, user, or some combination) in cases of highly autonomous AI systems could hinder the supervisory task of ensuring firms maintain adequate control and responsibility over their AI systems.

The limited explainability of advanced AI models discussed above can also make it harder to detect inappropriate use of data or use of unsuitable data in AI-based applications in finance. AI models are highly reliant on the quality, adequacy and completeness of the data they are trained on: biases or errors in the data can lead to biased or discriminatory outcomes for financial consumers, at times unintentionally (e.g. creditworthiness assessment) (OECD, 2021[2]). Identifying algorithmic bias is crucial to ensuring fair and

ethical use of AI in finance but can be obstructed by the lack of explainability discussed above, which may hinder the ability of supervisors to identify and mitigate bias. Assessing data governance practices within firms is an increasingly important supervisory task that could be challenged by the combination of model opacity and potential difficulties to verify the provenance, accuracy, completeness, relevance, and appropriateness of the vast amount of unstructured data used by advanced AI models. This extends beyond raw data to include metadata, documentation of data processing pipelines and data governance policies. The use of proprietary datasets, or the unwillingness of supervised entities to disclose detailed data-processing methodologies due to commercial sensitivity or contractual obligations where these are provided by third-party vendors, can exacerbate these challenges.

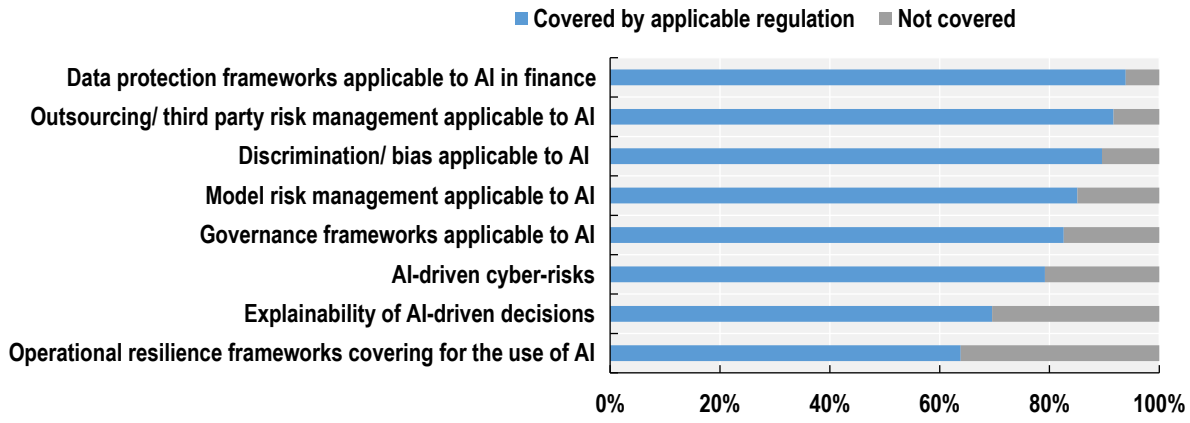
A final major governance issue lies in determining who instructs the AI model and what objectives are set. For example, in the use of AI for consumer finance, this may involve ensuring that the model is instructed to assist customers, instead of tailoring systems to maximise revenue from consumers. Governance structures should seek to address such concerns in their design by ensuring transparency in model objectives and fostering accountability for outcomes.

3 Supervisory practices to balance innovation and stability

The supervisory approach of OECD countries for the use of AI in finance leverages existing frameworks and practices which, although not AI-specific, are highly relevant and applicable given the principles-based, technology-agnostic approach to financial supervision. This also includes non-mandatory frameworks and high-level principles (e.g. those requiring firms to act with skill, care and due diligence and maintain adequate management and control) which remain applicable to AI governance in finance.

Most of the areas where challenges to AI supervision have been identified are also reported to be appropriately covered by regulatory frameworks in place in the vast majority of OECD countries (Figure 3.1). Therefore, possible challenges discussed therein lie in the interpretation of these rules at the practical implementation level, rather than in identified regulatory gaps. Several possible practices are being discussed in this Section to assist financial supervisors in overcoming such challenges, where these arise, and in ensuring effective oversight of AI activity by assisting supervisors to operationalise applicable rules.

Figure 3.1. Appropriate AI regulation is reported to be in place to address areas related to supervisory challenges



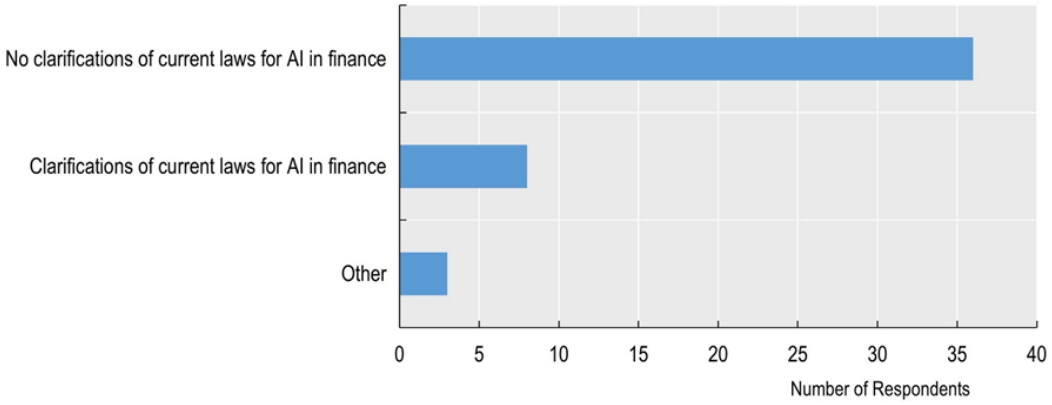
Note: Based on a total of 49 responding jurisdictions.
Source: 2024 OECD Survey on Regulatory Approaches to AI in Finance.

3.1. Consider carefully calibrated additional guidance on supervisory expectations/ supervisory guidance when this is warranted

The principles-based approach adopted by most OECD countries involves setting high-level outcomes or objectives that firms are expected to achieve, granting them flexibility in determining how to meet these standards and requirements. The primary advantage of this approach is its flexibility and adaptability as it can accommodate rapid AI innovation without requiring constant rule updates that may quickly become obsolete. However, principles can be perceived as ambiguous when it comes to highly complex AI systems, potentially leading to uncertainty for firms who report lack of regulatory clarity in certain markets. This could be due to potential difficulties that certain supervised entities encounter in interpreting and operationalising principles-based rules within the complex and evolving context of advanced AI systems, potentially giving rise to compliance gaps in certain markets.

Where supervised entities are facing challenges due to such perceived lack of clarity, additional guidance as to the interpretation of high-level principles could be beneficial in providing legal certainty for firms and promoting consistent regulatory outcomes. Some supervisory authorities have indeed begun articulating explicit supervisory expectations for the oversight of AI application in financial services, through both internal and public-facing guidance. Examples of such guidance provided in this note are associated primarily with model risk management, given the central role of such frameworks. However, clarifications have been issued only by a small number of financial regulators or supervisors, despite unique issues arising in the deployment of AI innovation and potential associated challenges discussed in this note (OECD, 2024[1]) (Figure 3.2).

Figure 3.2. Clarifications around the applicability of existing rules/ regulations/ other policy frameworks on AI applications in finance



Note: Based on a total of 49 responding jurisdictions.
Source: OECD (2024[1]), Regulatory approaches to Artificial Intelligence in finance, https://www.oecd.org/en/publications/regulatory-approaches-to-artificial-intelligence-in-finance_f1498c02-en.html

Additional guidance and clarification setting expectations for the industry could therefore be beneficial in jurisdictions where supervised entities are reporting challenges due to perceived lack of clarity to assist firms in their compliance in jurisdictions where supervised entities are reporting facing challenges due to perceived lack of clarity. This could include areas such as model risk management requirements; explainability of models; assessment of robustness of output and validation, including regarding ethical considerations (e.g., fairness); governance of AI systems, including data management. Eliminating any perceived ambiguity in these areas may also enhance legal certainty, thereby fostering greater confidence in the regulatory environment and encouraging further investment in, and broader responsible adoption of

AI in finance. Clarifying these requirements would also strengthen the ability of supervisory authorities to resolve potential ambiguities at the internal supervisory level, should these arise, thereby strengthening their capacity to discharge their oversight responsibilities more effectively. Any guidance provided should be very carefully designed and calibrated, to avoid negatively affecting AI adoption by impeding firms' ability to flexibly explore new technologies. Overly prescriptive approaches should be avoided, given the rapid pace of technological innovation, and a risk-based approach may be most effective in enabling financial institutions to address key risks where warranted.

Supervisory initiatives could be encouraged to promote convergence in the interpretation and application of existing rules, both at the national/regional and at the international levels. In jurisdictions where dedicated regulations for AI have been introduced, such convergence entails the systematic identification and resolution of any overlaps or inconsistencies between these new AI-specific rules and pre-existing sectorial rules. In such cases, consideration can also be given to the possible integration of such newly introduced requirements into existing sectorial frameworks to support the simplification and streamlining of both the oversight activity and the compliance efforts of supervised entities. Additionally, international policy dialogue and coordination, information sharing, and efforts to support greater alignment between domestic and international regulators could be beneficial in effectively supporting wider responsible deployment of AI in finance.

3.2. Encourage public-private cooperation, leveraging sandboxes and novel AI model testing

Direct supervisory engagement with the industry, an essential pillar of effective oversight, serves as a vital channel for dialogue with entities deploying AI systems, one that warrants further reinforcement and emphasis. Such interaction deepens supervisory understanding of the practical deployment of AI innovation and operational contexts of such technologies. Importantly, it also helps enhance the capacity of authorities to identify and address emerging risks in a timely and well-informed manner.

Close and sustained engagement with the industry can yield significant benefits for supervised entities, improving the authorities' understanding of any challenges encountered by supervised firms in their compliance efforts. This is particularly pertinent when it comes to bridging any gap between principles-based financial rules and their operationalisation, especially in light of the growing complexity and distinctive characteristics of advanced AI systems, as discussed in this note.

Enhanced ways of proactive engagement between supervisors and industry stakeholders beyond the standard supervisory activities could also be considered as a way to foster mutual understanding. Building on traditional supervisory activities such as on-site inspections, thematic assessments, systematic data gathering, all of which support greater mutual understanding, authorities can further engage with the industry to ensure that financial institutions remain compliant with regulatory standards, adequately manage risk, and uphold market integrity.

Innovation facilitators focusing on AI applications in finance, such as regulatory sandboxes, enable firms to test emerging technologies in a controlled environment under the direct supervision of authorities, providing a structured setting in which legal and operational challenges can be identified and addressed at an early stage (OECD, 2025[41]; 2023[42]). The usefulness of experimentation through sandboxes has also been acknowledged in the revised OECD AI Principles. Such arrangements, when carefully designed, can address compliance uncertainties when it comes to the deployment of AI models, while also promoting a culture of experimentation critical to the wider diffusion of AI innovation. Such collaborative engagement can contribute to a better understanding of the impact of AI on financial service provision, on measurable benefits and ensuing risks, and can inform the development of more adaptive and anticipatory supervisory approaches, ultimately reinforcing both innovation ecosystems and the integrity of financial markets.

More novel initiatives involving model testing can cultivate productive dialogue between firms developing or deploying AI models and supervisory bodies, fostering mutual understanding and supporting model validation. The UK FCA AI Live Testing initiative is an example of such novel approach in providing tailored support to financial firms, including around the exploration of output-driven model assessment and validation methods (FCA, 2025[43])(see Box 3.1). In Japan, in June, the JFSA launched a Public-Private AI Forum to deepen discussions on key issues, such as personal data protection, cybersecurity, model risk management, misuse of AI for financial crimes, and talent development (FSA Japan, 2025[44]). These initiatives encourage a culture of responsible innovation by aligning supervisory expectations with industry practices. The collaborative nature of testing can also help bridge knowledge gaps and foster trust between regulators and AI developers. Ultimately, such efforts have the potential to contribute to more robust and transparent AI governance ecosystems across the financial sector.

Box 3.1. AI Live Testing by the UK Financial Conduct Authority (FCA)

The UK FCA is planning to launch AI Live Testing, as part of the existing AI Lab, to support firms' safe and responsible deployment of AI and achieve positive outcomes for UK consumers and markets. AI Live Testing closely aligns with other innovation approaches that form part of the AI Lab, including the Supercharged Sandbox and the Digital Sandbox.

The UK's principles-based and outcomes-focused regulatory framework aims to afford firms the flexibility to innovate in the delivery of financial services. UK authorities have clarified that they intend to avoid introducing additional AI-specific regulation, opting instead to rely on existing frameworks. Through AI Live Testing, firms will have the opportunity to trial AI models in real-world conditions, thereby building confidence in their performance while receiving regulatory guidance and assurance to support appropriate deployment. This includes the exploration of output-driven validation methodologies to assess model performance, providing regulatory support and regulatory comfort to support appropriate deployment.

The aim is to instil confidence and certainty among firms to invest in AI technologies in ways that foster growth and deliver positive outcomes for consumers and markets. Concurrently, the regulatory authorities seek to collaborate with industry to deepen understanding of AI-related risks and to explore effective mitigation strategies, thereby reinforcing the authorities' commitment to safe and responsible AI development. Some of the questions the FCA aims to explore include:

- what input-output validation may be needed to build confidence that AI-generated outcomes are likely to meet regulatory expectations;
- how to assess if an AI model is sufficiently robust, including the choice of relevant metrics against which to assess robustness, and how firms should monitor robustness;
- if the output of an AI model is not explainable and/or not sufficiently robust, what are the implications for its potential use in UK financial markets;
- how to measure the degree of bias in an AI model and, by implication, what degree of debiasing may be appropriate for any given use case;
- how consumer groups, including vulnerable consumers, may be impacted;
- what processes are in place to address poor/unintended AI model outcomes when they arise.

Source: FCA (2025[43]), Engagement Paper: Proposal for AI Live Testing, https://www.fca.org.uk/publication/call-for-input/ai-testing-pilot-engagement-paper.pdf

Emphasis on the multidisciplinary approach to public-private engagement is also warranted in this domain, in line with the OECD AI Principles (OECD, 2019[45]; 2024[46]). This can include, for example, collaboration with AI infrastructure and compute providers given the importance of such third-party service providers in AI development; increased deployment of AI innovation in finance has been achieved thanks to such vendors and their beneficial role in wider diffusion of AI in the sector cannot be disregarded.

The Supercharged Sandbox of the UK FCA, part of the AI Lab, is an example of such multi-stakeholder engagement for testing purposes. Participating firms in this innovation facilitator are granted access to greater computing capabilities, enhanced datasets, and more advanced AI-related tooling (FCA, 2024[47]). This can help accelerate experimentation with AI models in a safe and supervised environment, enabling firms to better understand regulatory expectations. Access to computing helps lower the barriers to entry for smaller firms and has the potential to foster a more inclusive innovation ecosystem. This can help firms

develop and validate AI solutions more efficiently, while allowing regulators to observe emerging risks and practices in real time, helping to align innovation with public policy objectives.

The abovementioned examples underscore the indispensable role of supervisory expertise, judgement, and institutional capacity as prerequisites for the effective implementation of such efforts. Strengthening these capabilities in the context of novel approaches to public-private engagement, such as the examples mentioned above, may be further supported through collaboration with technical experts from academia and the private sector, thereby advancing a multidisciplinary approach to AI governance, as advocated in the OECD AI Principles.

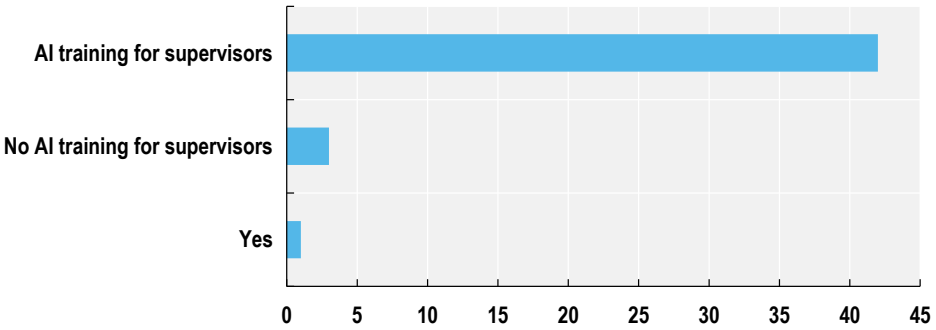
3.3. Invest in supervisory capacity, upskilling and use of AI-driven SupTech

The need to equip policy makers, especially financial supervisors, with the right tools and skills for effective AI oversight in finance is widely acknowledged (OECD-FSB, 2024[48]).The specificities of AI innovation warrant the existence of well-resourced supervisory teams with deeper domain and technical expertise and a comprehensive understanding of both sectorial and technical implications of AI innovation in finance. Sufficient resources are required to effectively oversee and continuously monitor the evolution of AI deployment in finance, and to allow supervisors to keep abreast of the rapid advances at the technological front. Without deep understanding and continuous monitoring of AI developments, supervisors will be ill-equipped to effectively assess compliance with principles underpinning financial regulation.

Increased capacity and upskilling of financial supervisors will therefore be necessary to achieve monitoring and oversight objectives, but also to enable authorities to develop and deploy AI as part of the supervisory activity. In addition to attracting talent with expertise in AI-related topics, it will be important to train and upskill existing teams allowing them to combine their domain-specific expertise with a deeper technical understanding of AI systems. Indeed, most jurisdictions responding to the 2024 OECD survey have ongoing training and other activities to promote the upskilling of supervisors (Figure 3.3).

Upskilling and capacity building efforts need to be sustained and continuous, rather than ad hoc or one-off initiatives. Given the exceptionally rapid innovation cycles involved in the AI field, supervisors face the constant challenge of keeping their knowledge, skills, and oversight frameworks current with the technological frontier. Investment is also required to conduct further research into the potential long-term impacts of AI on financial market structures, competition, and financial stability.

Figure 3.3. Action to promote the upskilling of supervisors in relation to AI ongoing in majority of OECD countries



Note: Based on a total of 49 responding jurisdictions.
Source: 2024 OECD Survey on Regulatory Approaches to AI in Finance.

The deployment of AI-enabled Supervisory Technology (SupTech) tools can support more effective oversight of financial activity while offering supervisors direct exposure to AI innovation, thereby deepening their practical understanding of AI systems. SupTech benefits are widely acknowledged and documented by financial authorities and experimentation efforts are underway both at the national and at the international levels (FSB, 2020[49]). SupTech tools based on AI can enhance monitoring functions (e.g. assist large-scale data analysis, market surveillance and real-time monitoring, automated compliance verification, AI model monitoring, among other tasks). As AI is reshaping financial market dynamics, supervisors may need to increasingly integrate AI innovation into their own toolkits and supervisory responses to be able to effectively monitor and intervene in increasingly automated markets, enhance their responsiveness but also their foresight and capacity to anticipate emerging vulnerabilities.

At the operational level, coordinated efforts among supervisory authorities could enable the strategic pooling of expertise and institutional capacity when it comes to AI-based SupTech tools. The development or acquisition of SupTech applications involving AI can necessitate significant financial investment, robust technological infrastructure, and specialised internal expertise. These requirements may pose challenges for those supervisory authorities that may face resource constraints. International collaboration could offer a path to pool resources, share knowledge, and develop common tools. Joint engagements at the cross-border level for the development and sharing of SupTech solutions, or collaborative training initiatives, can also be beneficial. A collective approach can reduce duplication of efforts and strengthen the overall resilience and adaptability of supervisory regimes in the face of rapid technological change through the sharing of best practices (e.g. BIS Innovation Hub Project Aurora testing the use of AI for AML; ECB projects) (see Box 3.2).

Box 3.2. AI tools at the European Central Bank's (ECB) SupTech Hub

The ECB has established a dedicated SupTech Hub and developed numerous applications, including using AI innovation, promoting the dual role of supervisors as both overseers and users of AI.

The ECB is actively exploring over 40 potential GenAI use cases to support supervisors' daily tasks. Notable examples of such tools include 'Athena' for translating and analysing supervisory documents using textual analysis; 'Agora' which allows supervisors to query data lakes using natural language translated into code by AI; a 'Virtual Lab' cloud platform for ML development and collaboration; as well as 'Medusa' which aims to use AI to assist in drafting and benchmarking supervisory findings and 'Heimdall' tool that uses machine reading to automatically undertake fit and proper procedures.

Additionally, project Delphi uses natural language processing (NLP) to integrate market risk-based indicators and information from news items into a single web-based platform with a user-friendly interface for banking supervisors.

Source: ECB SSM website; Frederik Hoppe (2025[50]), Benefits from advanced technology infrastructure in supervision, https://www.bankingsupervision.europa.eu/press/supervisory-newsletters/newsletter/2025/html/ssm.nl250514.en.html; McCaul (2024[51]), The future of European banking supervision - connecting people and technology, https://www.bankingsupervision.europa.eu/press/speeches/date/2024/html/ssm.sp240918~522b3441ba.en.html

3.4. Encourage policymaker coordination across sectors and jurisdictions

Recognising the cross-cutting nature of AI, implicating authorities beyond the financial sector, as well as coordination and information sharing are vital for the safe adoption, development and deployment of AI. Policy alignment between different areas of policy making applicable to the use of AI in finance, such as

data policy, should also be encouraged at the supervisory level. As AI innovation is increasingly seen as a GPT, it warrants the joining up of authorities across sectors too.

As AI systems become more integrated into financial services, issues such as explainability, ethical use, and systemic impact could become more pressing, demanding increasing attention by supervisory authorities. Given the global nature of both AI innovation and financial market activity, collaborative approaches by supervisors across borders could also be beneficial to foster responsible innovation, while allowing for national specificities and priorities. International collaboration and information sharing can support the identification of emerging vulnerabilities (e.g. efforts around incident reporting (OECD, 2025[52])). Conversely, marked divergences in supervisory approaches and practices across jurisdictions may risk undermining the confidence of globally active financial markets participants, potentially discouraging broader investment in the development and deployment of AI.

3.5. Evolution of AI Supervision in finance: pushing the boundaries of tech neutrality?

While AI regulation across OECD economies follows a tech-neutral principles-based approach, translating and implementing these rules at the operational level may indicate a need to enrich supervisory practices and tools to account for AI specificities in a forward-looking approach. The distinctiveness of AI and its novelty compared to other technological innovation applied in finance, the pervasiveness of its impact on the real economy and across sectors, the speed at which it evolves, and the challenges raised in this report highlight that AI could be pushing the boundaries of tech-neutrality when it comes to the operationalisation of the applicable policy frameworks at the oversight level.

While existing, principles-based technology-neutral rules remain at the core of the supervisory exercise, the need to gain assurance that firms are adequately managing the inherent risks of AI (e.g. explainability, bias/fairness, robustness) may compel supervisors to delve into technology-specific methodologies or even metrics in the future. This could involve potentially considering methodologies or metrics for assessing acceptable levels of robustness, explainability, or fairness levels in particular cases; methodologies and techniques evaluating output robustness; the adequacy of explainability tools used, or tools to ensure unbiased outputs of AI-driven models, depending on the context and the use-case involved.

Maintaining a flexible and adaptive stance when it comes to the financial oversight activity at the practical level could help address reported supervisory challenges while allowing oversight to keep pace with technological advances. Allowing for technology-specific guidance or consideration of novel methodologies and techniques to enrich the supervisory toolkit could support supervisors in achieving the delicate balance between fostering innovation and ensuring stability. Consideration of new supervisory methods, techniques, and tools should take place in the context of continuous public-private dialogue. Proactive engagement with the finance industry in AI-specific testing through sandboxes or model testing, for example, can provide the confidence and clarity needed to encourage innovation while protecting markets and their participants, and safeguarding stability.

Continuous assessment of the existing supervisory landscape will be essential to ensure it remains fit for purpose, along with a willingness to enhance and adapt the supervisory tools to the evolving realities and nuances of AI innovation. Strengthening the capacity of supervisory authorities to address potential ambiguities can help unlock the finance sector's ability to further invest in productive AI innovation in a responsible manner. While regulation often targets entities rather than specific technologies, continued public-private dialogue may be valuable for exploring and refining supervisory methods and tools in the context of rapid technological advancement.

References

ACPR (2020), *Governance of Artificial Intelligence in Finance*, https://acpr.banque-france.fr/en/publications-and-statistics/publications/governance-artificial-intelligence-finance (accessed on 6 April 2021). [30]

ACPR Banque de France (2025), *La motivation du conseil par les robo-advisors : vers un éclairage apporté aux clients ? | Autorité de contrôle prudentiel et de résolution*, https://acpr.banque-france.fr/en/node/30113 (accessed on 14 June 2025). [40]

ASIC (2024), *Introducing mandatory guardrails for AI in high-risk settings: proposals paper*, https://consult.industry.gov.au/ai-mandatory-guardrails/submission/view/174 (accessed on 10 December 2025). [9]

BaFin (2023), *Circular 05/2023 (BA) - Minimum Requirements for Risk Management …*, https://www.bafin.de/SharedDocs/Downloads/EN/Rundschreiben/dl_rs_0523_marisk_ba_en.html (accessed on 2 July 2025). [22]

Bank of England (2025), *Financial Stability in Focus: Artificial intelligence in the financial system | Bank of England*, https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025 (accessed on 11 July 2025). [14]

Bank of England (2023), *SS1/23 – Model risk management principles for banks | Bank of England*, https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss (accessed on 30 April 2024). [20]

BIS FSI (2024), "FSI Insights on policy implementation No 63 Regulating AI in the financial sector: recent developments and main challenges", http://www.bis.org/emailalerts.htm. (accessed on 3 July 2025). [55]

Central Bank of Ireland (2023), "Data Ethics Within Insurance", https://www.centralbank.ie/docs/default-source/regulation/industry-market-sectors/insurance-reinsurance/solvency-ii/communications/data-ethics-within-insurance.pdf?sfvrsn=54219f1d_4 (accessed on 24 April 2024). [6]

Daníelsson, J., R. Macrae and A. Uthemann (2022), "Artificial intelligence and systemic risk", *Journal of Banking & Finance*, Vol. 140, p. 106290, https://doi.org/10.1016/J.JBANKFIN.2021.106290. [17]

Deutsche Bundesbank and BaFin (2021), “Machine learning in risk models-Machine learning in risk models-Characteristics and supervisory priorities”, https://www.bafin.de/SharedDocs/Veroeffentlichungen/EN/BaFinPerspektiven/2019_01/bp_19-1_Beitrag_SR3_en.html (accessed on 26 April 2024). [21]

EBA (2023), “Machine Learning for IRB Models”, https://www.eba.europa.eu/sites/default/files/document_library/Publications/Reports/2023/1061483/Follow-up%20report%20on%20machine%20learning%20for%20IRB%20models.pdf (accessed on 2 July 2025). [27]

ECB (2024), “ECB guide to internal models”, https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guides202402_internalmodels.en.pdf (accessed on 2 July 2025). [28]

EIOPA (2024), *Factsheet on the regulatory framework applicable to AI systems in the insurance sector - EIOPA*, https://www.eiopa.europa.eu/publications/factsheet-regulatory-framework-applicable-ai-systems-insurance-sector_en (accessed on 30 June 2025). [8]

EIOPA (2021), “Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the european insurance sector - report from EIOPA´s Consultative Expert Group on Digital Ethics in insurance”, https://doi.org/10.2854/49874. [37]

ESMA (2024), *ESMA provides guidance to firms using artificial intelligence in investment services*, https://www.esma.europa.eu/press-news/esma-news/esma-provides-guidance-firms-using-artificial-intelligence-investment-services (accessed on 19 September 2024). [7]

EU (2024), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (accessed on 9 January 2025). [5]

EU (2022), *Directive (EU) 2022/2556 of the European Parliament and of the Council of 14 December 2022 amending Directives 2009/65/EC, 2009/138/EC, 2011/61/EU, 2013/36/EU, 2014/59/EU, 2014/65/EU, (EU) 2015/2366 and (EU) 2016/2341 as regards digital operational resilience for the financial sector [DORA Directive]*, https://eur-lex.europa.eu/eli/dir/2022/2556/oj/eng (accessed on 20 May 2025). [19]

EU (2018), *General Data Protection Regulation (GDPR) – Legal Text*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 (accessed on 20 January 2025). [56]

EU (2014), *Markets in Financial Instruments Directive*, Official Journal of the European Union, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065 (accessed on 22 November 2023). [18]

FCA (2025), *Engagement Paper: Proposal for AI Live Testing*, https://www.fca.org.uk/publication/call-for-input/ai-testing-pilot-engagement-paper.pdf (accessed on 17 June 2025). [43]

FCA (2024), *AI Lab | FCA*, https://www.fca.org.uk/firms/innovation/ai-lab#section-supercharged-sandbox (accessed on 4 July 2025). [47]

FDIC (2017), *Adoption of Supervisory Guidance on Model Risk Management | FDIC*, https://www.fdic.gov/news/financial-institution-letters/2017/fil17022.html (accessed on 24 April 2024). [25]

Federal Reserve (2011), *The Fed - Supervisory Letter SR 11-7 on guidance on Model Risk Management -- April 4, 2011*, https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm (accessed on 18 March 2021). [23]

FFIEC (2009), *Interagency Fair Lending Examination Procedures*, https://www.ffiec.gov/PDF/fairlend.pdf (accessed on 24 April 2024). [4]

Filippucci, F. et al. (2024), "The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges", *OECD Artificial Intelligence Papers*, No. 15, OECD Publishing, Paris, https://doi.org/10.1787/8d900037-en. [15]

Filippucci, F., P. Gal and M. Schief (2024), "Miracle or Myth? Assessing the macroeconomic productivity gains from Artificial Intelligence", *OECD Artificial Intelligence Papers*, No. 29, OECD Publishing, Paris, https://doi.org/10.1787/b524a072-en. [16]

FINMA (2024), "FINMA Risk Monitor 2024 2". [13]

FINMA (2023), *FINMA Risk Monitor*, https://www.finma.ch/en/~/media/finma/dokumente/dokumentencenter/myfinma/finma-publikationen/risikomonitor/20231109-finma-risikomonitor-2023.pdf?sc_lang=en&hash=27012208F65AFDAC00DBE199ADF8DB40 (accessed on 14 June 2025). [26]

Frederik Hoppe (2025), *Benefits from advanced technology infrastructure in supervision*, https://www.bankingsupervision.europa.eu/press/supervisory-newsletters/newsletter/2025/html/ssm.nl250514.en.html (accessed on 4 July 2025). [50]

FSA Japan (2025), "AI Discussion Paper Version 1.0 Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector Overview". [32]

FSA Japan (2025), *Public-private AI Forum「金融庁AI官民フォーラム」（第２回）議事要旨：金融庁*, https://www.fsa.go.jp/singi/ai_forum/gijiyoshi/20251024.html (accessed on 10 December 2025). [44]

FSA Japan (2021), "Principles for Model Risk Management", (accessed on 30 June 2025), https://www.fsa.go.jp/common/law/ginkou/pdf_03.pdf (accessed on 3 July 2025). [31]

FSB (2025), "Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector", http://www.fsb.org/emailalert (accessed on 27 October 2025). [11]

FSB (2024), *The Financial Stability Implications of Artificial Intelligence - Financial Stability Board*, https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/ (accessed on 30 June 2025). [10]

FSB (2023), *Final Report on Enhancing Third-party Risk Management and Oversight – A Toolkit for Financial Institutions and Financial Authorities - Financial Stability Board*, https://www.fsb.org/2023/12/final-report-on-enhancing-third-party-risk-management-and-oversight-a-toolkit-for-financial-institutions-and-financial-authorities/ (accessed on 30 June 2025). [12]

FSB (2020), *The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market developments and financial stability implications*, http://www.fsb.org/emailalert (accessed on 18 March 2021). [49]

FSC Korea (2021), *Guideline on the use of Artificial Intelligence in Financial Services*, https://www.fsc.go.kr/eng/pr010101/76209 (accessed on 3 July 2025). [33]

IBM (2025), *What Is Agentic AI?*, https://www.ibm.com/think/topics/agentic-ai (accessed on 14 June 2025). [53]

Ji, Z. et al. (2023), "Survey of Hallucination in Natural Language Generation", *ACM Computing Surveys*, Vol. 55/12, https://doi.org/10.1145/3571730. [54]

MAS (2024), "Artificial Intelligence Model Risk Management - Observations from a thematic review - Formation Paper", *Artificial Intelligence Model Risk Management | 2*. [35]

MAS (2018), "Monetary Authority Of Singapore Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector Principles to Promote FEAT in the Use of AI and Data Analytics in Singapore's Financial Sector Monetary Authority of Singapore", https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data (accessed on 6 December 2023). [36]

McCaul, E. (2024), "The future of European banking supervision - connecting people and technology", https://www.bis.org/review/r240924i.htm (accessed on 4 July 2025). [51]

NAIC (2023), "NAIC Model Bulletin: Use of Artificial Intelligence systems by Insurers", https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf.pdf (accessed on 2 May 2024). [39]

NIST (2023), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)", https://doi.org/10.6028/NIST.AI.100-1. [34]

OCC (2011), *Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management | OCC*, https://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-12.html (accessed on 24 April 2024). [24]

OECD (2025), *Asia Capital Markets Report 2025 - Chapter 5: AI innovation facilitators in Asia*, https://www.oecd.org/en/publications/2025/06/asia-capital-markets-report-2025_8c82611c.html (accessed on 4 July 2025). [41]

OECD (2025), "Towards a common reporting framework for AI incidents", *OECD Artificial Intelligence Papers*, No. 34, OECD Publishing, Paris, https://doi.org/10.1787/f326d4ac-en. [52]

OECD (2024), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 (accessed on 22 November 2024). [46]

OECD (2024), *Regulatory approaches to Artificial Intelligence in finance*, OECD Publishing, Paris, https://www.oecd.org/en/publications/regulatory-approaches-to-artificial-intelligence-in-finance_f1498c02-en.html (accessed on 17 September 2024). [1]

OECD (2023), "Generative artificial intelligence in finance", *OECD Artificial Intelligence Papers*, No. 9, OECD Publishing, Paris, https://doi.org/10.1787/ac7149cc-en. [3]

OECD (2023), *Supporting FinTech Innovation in the Czech Republic: Regulatory Sandbox Design Considerations | en | OECD*, https://www.oecd.org/publications/supporting-fintech-innovation-in-the-czech-republic-081a005c-en.htm (accessed on 12 July 2023). [42]

OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges and Implications for Policy Makers*, https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf (accessed on 14 September 2021). [2]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en. [45]

OECD-FSB (2024), "OECD-FSB Roundtable on Artificial Intelligence (AI) in Finance", https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/digital-finance/OECD%20%E2%80%93%20FSB%20Roundtable%20on%20Artificial%20Intelligence%20(AI)%20in%20Finance.pdf (accessed on 4 July 2025). [48]

OSFI (2024), *Draft Guideline E-23 – Model Risk Management - Office of the Superintendent of Financial Institutions*, https://www.osfi-bsif.gc.ca/en/guidance/guidance-library/draft-guideline-e-23-model-risk-management (accessed on 30 April 2024). [38]

OSFI (2024), *OSFI updates model risk management guidance and launches public consultation - Office of the Superintendent of Financial Institutions*, https://www.osfi-bsif.gc.ca/en/news/osfi-updates-model-risk-management-guidance-launches-public-consultation (accessed on 30 April 2024). [29]

Notes

[1] Sandboxes, in this context, are testing environments whose primary purpose is to allow experimentation and evaluation of AI models in a controlled setting before broader deployment.

[2] Credit scoring or risk assessment and pricing in relation to natural persons in the case of life and health insurance.

[3] See for example FEAT principles, Model AI Governance Framework collaboration.

[4] See (IBM, 2025[53]): Unlike traditional AI models, which operate within predefined constraints and require human intervention, Agentic AI exhibits autonomy, goal-driven behaviour and adaptability. The term "Agentic" refers to these models' agency, or, their capacity to act independently and purposefully. Agentic AI builds on GenAI techniques by using large language models (LLMs) to function in dynamic environments. While generative models focus on creating content based on learned patterns, agentic AI extends this capability by applying generative outputs toward specific goals. A generative AI model like

OpenAI's ChatGPT might produce text, images or code, but an agentic AI system can use that generated content to complete complex tasks autonomously by calling external tools.

[5] Capable of learning and adapting their behaviour over time based on new data and input fed into the model, including user prompts in GenAI models, often in ways that are not fully predictable (OECD, 2023[3]).

[6] Artificial hallucination refers to the phenomenon of a machine, such as a chatbot, generating seemingly realistic sensory experiences that do not correspond to any real-world input. This can include visual, auditory, or other types of hallucinations. Artificial hallucination is not common in chatbots, as they are typically designed to respond based on pre-programmed rules and data sets rather than generating new information. However, there have been instances where advanced AI systems, such as generative models, have been found to produce hallucinations, particularly when trained on large amounts of unsupervised data (Ji et al., 2023[54]).

[7] Treating the AI models as though they have human-like qualities (BIS FSI, 2024[55]).

[8] Interpretability refers to the meaning of the model's output in the context of their designed functional purposes, while explainability refers to a representation of the mechanisms underlying the AI system's operation (NIST, 2023[34]).

[9] E.g. GDPR obligations in the EU: while the regulation does not explicitly use the term 'explainability', it implies a right to meaningful information in the context of automated decision-making, requiring that data subjects be informed about "the logic involved" in automated decisions (Article 15(1)(h)), and suggesting that individuals should receive an explanation of the decision reached after such processing (Recital 71) (EU, 2018[56]).