AIFS0030

Written evidence submitted by Techno-Regulatory AI Sandbox (TRAIS)

We are submitting this response to the UK Parliament's Treasury Committee inquiry about the use of AI in banking, pensions and other financial services,[1] on behalf of the Techno-Regulatory AI Sandbox (TRAIS) team, a Responsible AI UK[2] funded research project in partnership with the Centre for Responsible AI in India (CeRAI).[3] The TRAIS project is a multidisciplinary effort, involving legal scholars and technologists from the UK and India, which aims to develop responsible AI models compliant with tailored regulations within the banking and finance sector. We adopt a two-pronged approach: (i) creating evidence-based hypothetical regulatory requirements (both umbrella and sector-specific, addressing reliability, fairness, and privacy), and (ii) building AI models adhering to these hypotheticals while preserving utility. The project is an 18-month endeavour that officially started in January 2025 as a result of an International Sandpit.[4] Our response below provides evidence drawn from our academic research and literature review, and it aims to further inform Parliamentary decision making and discussions.

Executive Summary

Productivity Benefits [par. 1-8]

- AI in the UK financial services sector boosts productivity by automating tasks, improving decision-making, and enhancing customer experiences, with notable applications in risk management, customer service, underwriting, and fraud detection using machine learning (ML).
- Generative AI (GenAI) offers new opportunities, such as generating synthetic data to address data imbalance, identifying fraudulent identities, and improving operational efficiency in areas like data analysis, strategy formulation, client onboarding, and regulatory risk assessment.

Barriers to Adoption [par. 9-22]

- **Operational, technical, and security risks [par. 9-11]**: AI systems require continuous updates, particularly in fraud detection, leading to higher costs and complexity. AI models are also vulnerable to adversarial attacks and manipulation, raising concerns about system reliability and integrity.
- **Ethical and regulatory challenges [par. 12-15]:** Issues like data privacy, bias, transparency, must be addressed, particularly in light of data protection laws and, most recently, the EU AI Act. AI systems may unintentionally discriminate or undermine trust, especially when data collection practices or algorithmic decisions are unclear.

---

[1] See: Use of AI in banking, pensions and other financial services to be subject of new inquiry by MPs - Committees - UK Parliament
[2] See: Responsible AI
[3] See: Home | Centre for Responsible AI
[4] See: Responsible Ai International Sandpit – Responsible AI and https://cerai.substack.com/p/reflections-from-the-uk-rai-sandpit

AIFS0030

- **Trust [par 16-22]**: Resistance from both financial institutions and customers, driven by concerns over privacy, loss of human interaction, and trust in AI systems, can hinder adoption. Safeguards such as safety fine-tuning, alignment with human values, and internal red teaming are critical for responsible deployment.

<u>Regulatory Needs [par. 23-50]</u>

- **More and New Guidance [par. 23-37]**: Adapting existing regulations to emerging technology follows the international trend. However, more specificity and additional guidance is needed to address specific challenges imposed by AI. For example, how is a concept like explainability understood by regulators, by the developers and how could the resulting gap in understanding manifest in practice?
- **Transparency in Regulatory Action [par. 38-39]**: Regulators need to disclose aggregate data on negative impacts of technology and decision-making processes during regulatory actions, allowing further scrutiny by academia, and other stakeholders.
- **Regulatory Sandboxes [par. 40-41]**: These are positive initiatives, seen as good practices for identifying regulatory gaps in the context of changing technologies, and the agency should continue to invest its resources in such endeavours.
- **Government's Regulatory Role [par. 42-44]**: Using a hybrid approach, the government needs to provide support to various stakeholders in application of principles and development of responsible training models and AI applications.
- **Coordination between Agencies [45-48]**: Regulators must ensure consistency in guidance issued by them to avoid regulatory burden and increase compliance.
- **Implementing organisational measures [par 49]:** A good practice is to enhance data governance by implementing training, validation, testing, and documentation practices throughout the lifecycle of AI models, ensuring proper traceability, regulatory compliance, and incident logging.
- **Upskilling of Regulators [par. 50]**: Investment is required in training and upskilling its personnel and hiring additional resources where required.

A. AI Adoption in the Banking and Financial Services: Productivity Benefits

1.   **AI has the potential to significantly enhance productivity within the UK financial services sector, as is already being demonstrated across a range of disciplines**.[5] These examples highlight the instrumental role AI can play in driving efficiency and accelerating productivity – particularly in services that involve content creation or information processing.[6] Although the UK financial services sector is diverse – comprising banks, insurance companies, pension funds, and other financial institutions – AI has the capability to improve productivity

---

[5] J. Mao, B. Chen, and J. C. Liu, 'Generative Artificial Intelligence in Education and Its Implications for Assessment', *TechTrends*, vol. 68, no. 1, pp. 58–66, Jan. 2024, doi: 10.1007/s11528-023-00911-4;P. Zhang and M. N. Kamel Boulos, 'Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges', *Future Internet*, vol. 15, no. 9, p. 286, Aug. 2023, doi: 10.3390/fi15090286; H. Al Naqbi, Z. Bahroun, and V. Ahmed, 'Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review', *Sustainability*, vol. 16, no. 3, p. 1166, Jan. 2024, doi: 10.3390/su16031166
[6] M. R. Morris, 'Scientists' Perspectives on the Potential for Generative AI in their Fields', Apr. 04, 2023, *arXiv*: arXiv:2304.01420. doi: 10.48550/arXiv.2304.01420.

AIFS0030

across the board. It can automate routine tasks, enhance decision-making, and optimise customer experiences, delivering benefits throughout the entire industry.[7]

2. **While AI techniques have long been employed in the finance domain,[8] the latest advances in AI agents have significantly broadened their applicability and dramatically improved their utility**. For instance, consider risk management, where AI refines credit risk assessments and streamlines regulatory compliance;[9] customer service, where AI-powered chatbots and virtual assistants provide personalised banking experiences;[10] and loan and insurance underwriting by automating claims processes[11] and improving wealth management through robot-advisors.[12]

3. More specifically, AI and machine learning (ML) can help with fraud detection, by automatically flagging suspicious transactions in real time.[13] This would be most useful as UK Finance reported fraud related to cards totalled £524.5 million in 2021.[14] Fraud detection systems that include AI/ML techniques can simultaneously learn and train real and potential security threat responses while flagging anomalous behaviours for further investigation. Algorithmic training, back-testing, and validation using large datasets of historical debit or credit card transaction data help label events as either 'fraud' or 'non-fraud' and can then be stopped in real time.[15]

4. AI/ML algorithms, such as Bayes logistic regression, decision trees, random forests, support vector machines, and artificial neural networks, can reduce the number of transactions wrongly classified as 'false positives' and identify new trends and signals.[16] Some companies

---

[7] Balakrishna, S., Arulkumar, V., Srihari, M., & Rohith, C. (2023). Usage of Machine Learning and Artificial Intelligence in Industry 4.0 and Banking Sector. *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 1282–1287. https://doi.org/10.1109/ICSSIT55814.2023.10060952.

[8] K. Balaji, 'Revolutionizing High-Frequency Trading: The Impacts of Financial Technology and Data Science Innovations', in *Advances in Finance, Accounting, and Economics*, H. Chen, Ed., IGI Global, 2025, pp. 103–124. doi: 10.4018/979-8-3693-8186-1.ch004.

[9] Aziz, S., & Dowling, M. (2019). Machine Learning and AI for Risk Management. In T. Lynn, J.G. Mooney, P. Rosati, & M. Cummins (Eds.), *Disrupting Finance* (pp. 33–50). Springer International Publishing. https://doi.org/10.1007/978-3-030-02330-0_3; Yazdi, M., Zarei, E., Adumene, S., & Beheshti, A. (2024). Navigating the Power of Artificial Intelligence in Risk Management: A Comparative Analysis. *Safety*, *10*(2), 42.https://doi.org/10.3390/safety10020042

[10] Bisht, S., Sengupta, S., Tewari, I., Bisht, N., Pandey, K., & Upadhyay, A. (2024). AI-Driven Tools Transforming The Banking Landscape: Revolutionizing Finance. *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, *1*, 934–938. https://doi.org/10.1109/ICACCS60874.2024.10717099; Uprikar, V., Mahalle, A., Tighare, D., Kodmalwar, P., Landage, H., & Thakare, S. (2024). Assessing the adoption and impact of AI chatbots in customer service: A survey-based study in Indian service industries. *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*, 1–6. https://doi.org/10.1109/IDICAIEI61867.2024.10842925.

[11] Z. Bednarz and K. Manwaring, 'Hidden depths: The effects of extrinsic data collection on consumer insurance contracts', *Computer Law & Security Review*, vol. 45, p. 105667, Jul. 2022, doi: 10.1016/j.clsr.2022.105667; Jansen, M., Nguyen, H. Q., & Shams, A. (2025). Rise of the Machines: The Impact of Automated Underwriting. *Management Science*, *71*(2), 955–975. https://doi.org/10.1287/mnsc.2024.4986.

[12] Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. *Journal of Behavioral and Experimental Finance*, *27*, 100297. https://doi.org/10.1016/j.jbef.2020.100297.

[13] Buchanan, B. G., & Wright, D. (2021). The impact of machine learning on UK financial services. *Oxford Review of Economic Policy*, *37*(3), 537–563. https://doi.org/10.1093/oxrep/grab016.

[14] UK Finance, 'ANNUAL FRAUD REPORT: THE DEFINITIVE OVERVIEW OF PAYMENT INDUSTRY FRAUD IN 2021', 2022. Accessed: Mar. 13, 2025. [Online]. Available: https://www.ukfinance.org.uk/system/files/2022-06/Annual%20Fraud%20Report%202022_FINAL_.pdf.

[15] B. van Liebergen, 'Machine learning: A revolution in risk management and compliance?', *Journal of Financial Transformation*, vol. 45, pp. 60–67, 2017.

[16] M. Cecchini, H. Aytug, G. J. Koehler, and P. Pathak, 'Detecting Management Fraud in Public Companies', *Management Science*, vol. 56, no. 7, pp. 1146–1160, Jul. 2010, doi: 10.1287/mnsc.1100.1174.

AIFS0030

use audio tokenisation and text recognition to secure payments; others use textual analysis on unstructured data from the internet to identify security threats, including the dark web. However, a major challenge in validating these fraud detection methods is the lack of publicly available datasets, largely due to privacy and confidentiality concerns. In addition, fraud datasets are often highly imbalanced, with fraudulent transactions representing only a tiny fraction of total activity. This makes it difficult to train models that can accurately detect fraud without producing a high rate of false positives.

5.   **Continued investment into AI/ML algorithm development – combined with better access to high-quality, representative data – will be essential for advancing the effectiveness and reliability of fraud detection systems**.

6.   With the emergence of technologies such as large language models (LLMs), AI is increasingly being categorised into two broad groups: traditional AI and generative AI (GenAI). Traditional AI refers to the algorithms and ML models already widely deployed across the sector, while GenAI, which leverages LLMs and other generative models such as diffusion models, is an emerging area of innovation.[17]

7.   In fraud detection, and financial services in general, GenAI offers exciting possibilities. It can be used to generate synthetic data, helping to address the problem of data imbalance and improve model training. Furthermore, in cases of complex or sophisticated fraud, GenAI can play a critical role in detecting fake or synthetic identities – for example, by analysing document metadata to identify forged documentation or fraudulent submissions.[18]

8.   T**he usage of trained GenAI models is also likely to enhance operational efficiency across the financial service sector, especially in those areas that may benefit from the addition of a sizable number of skilled personnel in the workforce**. Illustrative examples of these include:

- The use of GenAI research agents to analyse the various aspects of data obtained from the market, customers, channels, products and services in the sector.
- The use of GenAI to contribute to the formulation of strategies and propositions as well as provide accelerated insight into the data and better coverage to data governance.
- The use of GenAI models to hyper-personalise communication so as to deliver brand marketing and sales while adhering to organisational policies.
- The use of virtual agents to automate end-to-end client experience and assist in channel engagement, helping clients on-board a new product/service segment by collecting personal information from them, assessing the risks associated with their profiles and helping them engage with the banks and other financial service providers; and
- The use of GenAI to enrich regulatory risk assessment including early warning systems for customers, adverse media screening, narrative engagement and other modes of combating financial crimes and frauds and providing cyber detection services.[19]

---

[17] IBM, 'Global Outlook for Banking and Financial Markets - Regenerate banking with AI', 2024.
[18] R. Bou Hamdan, Y. Elias, S. Madan, and D. Zhao, 'Navigating the dual nature of generative AI: enhancing fraud detection while mitigating risks', EY, Apr. 2024. Accessed: Mar. 14, 2025. [Online]. Available: https://www.ey.com/content/dam/ey-unified-site/ey-com/en-ca/industries/financial-services/documents/ey-how-generative-ai-can-be-employed-in-fraud-detection-and-prevention.pdf
[19] Deloitte, 'Changing the game: the impact of artificial intelligence on the banking and capital markets sector', June 2024. Accessed March 15, 2025. [Online] Available: https://www.deloitte.com/global/en/Industries/financial-

AIFS0030

B. Barriers to AI Adoption: Risks to Consumers and Financial Institutions

9.   What seems to be a common factor across these diverse use cases is the need for such GenAI models to collect, process and analyse bulk information from all the different stakeholders associated with the industry and use their analysis of such information as a key factor for subsequent partially or fully automated decision-making, which brings us to the next set of concerns as to whether there may be specific barriers to adoption, most notably significant ethical challenges and technical, operational and security-related risks.

10.   **Operational risks to consider stem from the need for AI systems to continually evolve, particularly in the rapidly changing landscape of fraud detection**. The constantly shifting tactics used by fraudsters require significant investment in updating and refining AI models to ensure they remain effective. **This ongoing need for adaptation introduces greater operational complexity and increases costs for financial institutions**.

11.   **Technical and security-related risks relate to AI models being inherently fragile and vulnerable to adversarial attacks[20]**. Techniques such as "jailbreaking" can exploit weaknesses in these systems, enabling bad actors to bypass controls or generate unintended outputs. There is also the risk that users may subtly and strategically alter their behaviour to manipulate AI decision-making, undermining the reliability and integrity of automated systems.

12.   Further, additional ethical challenges arise when considering the demands on privacy, bias, and the transparency of algorithmic decisions.

13.   **Privacy concerns are paramount as financial institutions collect, process, and analyse vast amounts of personal data to feed into AI systems**. Naturally, critical questions about the extent of data collection, consent, data protection, and the potential for surveillance and data breaches are raised.[21] Inherent biases in data or in the assumptions made during the development of AI can lead to discriminatory practices in lending, insurance underwriting, and customer service, disproportionately affecting marginalised groups.[22] And algorithmic transparency, or the lack thereof, complicates these issues further. Whether due to intellectual property issues or due to opaque decision-making processes of some AI models, this lack of transparency can undermine trust and accountability, making it difficult for customers and regulators to understand or challenge decisions made by AI systems.[23] This is made additionally

---

services/perspectives/changing-the-game.html. In particular, see the example of Natwest Bank in the UK NatWest having identified their usage of various AI-models as responsible for reducing fraud by 6% as a share of UK Industry, including a 90% reduction in
account opening fraud since 2019, an increase of five times in click-through for personalized lending on customized customer offers, that of Mastercard reducing credit card delinquency by 32% because of their AI use, as well as other instances of productivity increase in sectors such as insurance, banking and capital markets.

[20] El-Mhamdi, El-Mahdi, Sadegh Farhadkhani, Rachid Guerraoui, Nirupam Gupta, Lê-Nguyên Hoang, Rafael Pinot, Sébastien Rouault, and John Stephan. "On the impossible safety of large AI models." *arXiv preprint arXiv:2209.15259* (2022).

[21] H. K. Alhitmi, A. Mardiah, K. I. Al-Sulaiti, and J. Abbas, 'Data security and privacy concerns of AI-driven marketing in the context of economics and business field: an exploration into possible solutions', *Cogent Business & Management*, vol. 11, no. 1, p. 2393743, Dec. 2024, doi: 10.1080/23311975.2024.2393743.

[22] M. Tanna and W. Dunning, 'Bias and discrimination in the use of AI in the financial sector', in *Artificial Intelligence in Finance*, N. Remolina and A. Gurrea-Martinez, Eds., Edward Elgar Publishing, 2023, pp. 320–349. doi: 10.4337/9781803926179.00025.

[23] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck, 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges', *Philos. Technol.*, vol. 31, no. 4, pp. 611–627, Dec. 2018, doi: 10.1007/s13347-017-0279-x.

AIFS0030

challenging due to recent results demonstrating the impossibility of attributing decisions to features for sufficiently complex models[24], raising questions about the validity of the deployment of these systems in highly regulated industries such as finance.

14. Some regulatory challenges stem from legislation like the General Data Protection Regulation (GDPR), and, more recently, the EU AI Act. While GDPR provides a strong foundation for protecting individual privacy, it also imposes strict limitations on data sharing. Although legislation typically promotes user trust in systems, it still presents a dilemma for areas such as fraud detection, where access to shared data is integral to developing adaptive and anticipatory prevention mechanisms.

15. **A further challenge relates to resistance to change within traditional financial institutions.[25] Resistance driven by fear can arise at both operational and executive levels within financial institutions, particularly in response to structural changes across the industry. Additionally, there may be resistance from the customer base, often due to concerns about trust, privacy, and the perceived loss of human interaction**.[26]

16. These barriers indicate trust is a critical factor influencing the successful implementation of AI-driven tools;[27] without trust in the service-delivering tools, users may hesitate to adopt them, thus limiting their effectiveness.

17. While AI can reduce operational costs and enhance efficiency,[28] the adoption of generative AI in specific areas – such as customer service,[29] where it supports content creation, knowledge management, and information processing – presents relatively low risk, as long as the models are trained to refuse answering queries whose answers could damage the reputation of the financial institution, and the capabilities of the models used are limited.[30] In these contexts, GenAI could improve response times, personalise customer interactions, and manage large volumes of routine inquiries, allowing human agents to focus on more complex and high-value tasks.

18. Nonetheless, ensuring the safe and responsible deployment of GenAI remains essential, even in lower-risk applications. **Research suggests that safeguards such as safety fine-**

[24] Bilodeau, Blair, Natasha Jaques, Pang Wei Koh, and Been Kim. "Impossibility theorems for feature attribution." *Proceedings of the National Academy of Sciences* 121, no. 2 (2024): e2304406120.

[25] Rahman, M., Ming, T. H., Baigh, T. A., & Sarker, M. (2023). Adoption of artificial intelligence in banking services: An empirical analysis. *International Journal of Emerging Markets*, *18*(10), 4270–4300. https://doi.org/10.1108/IJOEM-06-2020-0724; A. A. Winecoff and E. A. Watkins, 'Artificial Concepts of Artificial Intelligence: Institutional Compliance and Resistance in AI Startups', in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, Oxford United Kingdom: ACM, Jul. 2022, pp. 788–799. doi: 10.1145/3514094.3534138.

[26] E. Jussupow, I. Benbasat, and A. Heinzl, *WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION*. 2020.

[27] Radhakrishnan, J., & Chattopadhyay, M. (2020). Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. In S. K. Sharma, Y. K. Dwivedi, B. Metri, & N. P. Rana (Eds.), *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation* (Vol. 617, pp. 89–99). Springer International Publishing. https://doi.org/10.1007/978-3-030-64849-7_9.

[28] Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893–1924. https://doi.org/10.1108/BPMJ-10-2019-0411.

[29] Huang, M.-H., & Rust, R. T. (2024). The Caring Machine: Feeling AI for Customer Care. *Journal of Marketing*, *88*(5), 1–23. https://doi.org/10.1177/00222429231224748.

[30] A. Kumar, A. Shankar, L. D. Hollebeek, A. Behl, and W. M. Lim, 'Generative artificial intelligence (GenAI) revolution: A deep dive into GenAI adoption', *Journal of Business Research*, vol. 189, p. 115160, Feb. 2025, doi: 10.1016/j.jbusres.2024.115160.

AIFS0030

tuning and alignment play a critical role in mitigating potential risks. [31] Safety fine-tuning involves refining GenAI models after their initial training to ensure they behave reliably and produce outputs that are appropriate for their intended use. Alignment ensures that the system's objectives and outputs are consistent with human values, organisational policies, and regulatory standards, which is particularly important in financial services where trust and compliance are paramount.[32] However, recent work indicates that safety alignment is a complex, multi-faceted problem that is domain-specific[33]. The creation of challenging benchmarks for the financial domain is thus of paramount importance.

19. **Another important safeguard is internal red-teaming, where dedicated teams stress-test AI systems by simulating adversarial attacks or misuse scenarios**. This proactive approach helps uncover potential vulnerabilities before they can be exploited or cause harm.[34] In addition, training models on diverse and representative datasets helps reduce biases and ensures the AI system can serve a wide range of customer needs fairly and accurately.

20. Privacy-preserving methods, including private training and secure data handling practices, are also essential to protect sensitive customer information and comply with data protection regulations such as GDPR.[35] **Beyond technical safeguards, concerns around financial literacy arise as AI-driven systems become more widespread**. Customers must understand how these systems work, the data used, and the implications of AI decisions. Financial institutions can bridge this gap by offering clear, accessible information, fostering trust and reducing the risks of misinterpretation.[36] By incorporating these safeguards, financial institutions can confidently deploy GenAI in customer-facing services, balancing efficiency gains with responsible risk management.

21. As an undesirable consequence, automating routine tasks may lead to job losses in areas such as data entry, back-office operations, and customer support. On the flip side, AI adoption can create new roles, including data scientists and AI specialists, underscoring the need for upskilling and addressing labour displacement. **Additionally, while AI is a powerful tool, it is not a magic solution but requires careful consideration for effective integration.**[37] Despite rapid advancements in AI, there is limited long-term evidence of its successful implementation, especially within financial services. If financial service providers view AI merely as a tool, they will recognise the importance of strategic planning, thoughtful use cases, and tactful adoption to ensure success.

---

[31] F. Barez and L. Marks, 'Safeguarding AI In Finance: Risks and Countermeasures', 2024. doi: 10.2139/ssrn.4937924.

[32] E. Firt, 'Calibrating machine behavior: a challenge for AI alignment', *Ethics Inf Technol*, vol. 25, no. 3, p. 42, Sep. 2023, doi: 10.1007/s10676-023-09716-8.

[33] Han, T., Kumar, A., Agarwal, C., & Lakkaraju, H. (2024). Medsafetybench: Evaluating and improving the medical safety of large language models. *arXiv preprint arXiv:2403.03744*.

[34] D. Ganguli *et al.*, 'Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned', 2022, *arXiv*. doi: 10.48550/ARXIV.2209.07858.

[35] R. Shokri and V. Shmatikov, 'Privacy-Preserving Deep Learning', in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver Colorado USA: ACM, Oct. 2015, pp. 1310–1321. doi: 10.1145/2810103.2813687.

[36] S. Vadari and C. Malladi, 'Generative Knowledge Management for Financial Inclusion Through Financial Literacy: A Systematic Review', *IUP Journal of Knowledge Management*, vol. 22, no. 1, pp. 39–75, Jan. 2024.

[37] M. Madanchian and H. Taherdoost, 'Barriers and Enablers of AI Adoption in Human Resource Management: A Critical Analysis of Organizational and Technological Factors', *Information*, vol. 16, no. 1, p. 51, Jan. 2025, doi: 10.3390/info16010051.

AIFS0030

22.   Ultimately, AI could improve productivity and efficiency in financial services, but it should not be viewed as the sole driver of these improvements, nor should it be adopted without careful considerations of potential harms to both institutions and customers.

C. Recommendations for Policymaking: Evaluating UK Regulatory Needs in the Banking and Financial Sector

a. Balancing Existing Regulations and New Guidance

23.   In the UK, the Financial Conduct Authority (FCA) and the Prudential Regulation Authority (PRA), both of which oversee banking and financial services, have distinct roles. The PRA ensures the financial stability of banks, while the FCA is focused on how banks engage with their clients and operate in financial markets.

24.   Following the release of the UK Government's AI White Paper (2023)[38] and a call for updates on AI strategies from regulatory agencies (2024),[39] both the FCA and the PRA outlined their responses, detailing their approach to regulating AI.[40] They emphasized a technology-neutral approach to supervision, meaning their principles, rules, and guidelines are applicable regardless of the technology employed by regulated entities. They have positioned themselves as an outcomes-focused regulator, emphasizing a shift from rigid compliance with detailed rules towards ensuring that firms achieve predefined policy objectives. This approach, which is increasingly seen as flexible, helps to focus oversight on desired outcomes rather than micromanaging the technologies used.[41]

25.   Such a regulatory stance has been advocated by market participants and is also aligned with international practices. Over time, financial authorities have shifted from a compliance-based model, focused on detailed rules, to one that prioritizes ensuring the sound operation of financial institutions. This approach is intended to safeguard financial stability, regardless of the specific technologies employed, including AI.

26.   Many stakeholders have argued that the risks posed by AI are not unique to the technology itself and can be addressed within existing regulatory frameworks. This technology-neutral approach has the potential to integrate AI into current financial regulation while balancing risk management with innovation. As OECD has previously pointed out (2021, 2023),[42] many of the risks related to AI are not necessarily new or unique to AI innovation but rather exacerbated and amplified when such innovation is used – or manifest in different ways[43].

---

[38] A pro-innovation approach to AI regulation - GOV.UK
[39] Regulators' strategic approaches to AI - GOV.UK
[40] Bank of England Response Letter & FCA Response Letter
[41] Siau, K. & Wang, W., 2020. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal of Database Management (JDM)*, 31(2), pp.14. DOI: 10.4018/JDM.2020040105.; Handrlica, J., Štěpán, V. & Nešpor, J., 2023. Forum shopping in regulatory sandboxes and the perils of experimental law-making. *Juridical Tribune - Review of Comparative and International Law*, 13(3), pp.408-426.
[42] OECD (2021), Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges and Implications for Policy Makers, https://www.oecd.org/en/publications/artificialintelligence-machine-learning-and-big-data-in-finance_98e761e7-en.html.; OECD (2023), Generative artificial intelligence in finance, https://www.oecd.org/fr/finances/generative-artificial-intelligence-in-finance-ac7149cc-en.htm.
[43] Preethi Seshadri, Sameer Singh, and Yanai Elazar. 2024. The Bias Amplification Paradox in Text-to-Image Generation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for

AIFS0030

27.   Globally, regulators have emphasized that existing financial regulations, such as those concerning consumer protection, model risk management, and cybersecurity, continue to apply irrespective of the technology used. The OECD (2024) highlights laws covering prudent business conduct, IT governance, and fairness remain relevant for all types of technologies, including AI.[44] This reflects a growing consensus that new technologies do not undermine existing safety standards but rather should be managed within established frameworks.

28.   The Financial Stability Institute (FSI) also acknowledges that many of the common concerns associated with AI, such as accountability, transparency, and data security, are not novel to the financial sector. These concerns have long been addressed through general financial regulations focused on risk management and consumer protection.[45]

29.   **While both the FCA and the PRA are examining how to adapt existing regulations to emerging technologies without creating new, AI-specific rules, their approach aligns with broader trends in the financial sector. However, as noted by the OECD, different jurisdictions may find it necessary to provide additional regulatory guidance tailored to the specific challenges presented by AI, particularly in areas where significant risks or customer impacts are involved**.[46]

30.   The FSI echoes this perspective, recognizing that AI presents unique challenges to the implementation of existing regulations. While most areas can be addressed through current frameworks, AI-specific guidance may be required in high-risk use cases. This suggests that regulators may need to clarify, revise, or even introduce new rules to manage the emerging risks posed by AI in financial services.[47]

31.   In the United States, similar discussions have taken place. The Department of Treasury (2024) highlighted that stakeholders have called for clearer guidance on ensuring compliance with consumer protection laws across both traditional and non-traditional financial players. Respondents suggested that such guidance could encourage more competition in the market while mitigating the risk of consumer harm. In response, the Treasury recommended further engagement with stakeholders to identify any gaps in existing regulations and explore solutions to address potential risks associated with AI.[48]

32.   **Key areas that may require additional regulatory guidance towards enforcement and interpretation of existing decisions and regulations include governance frameworks,**

Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 6367–6384, Mexico City, Mexico. Association for Computational Linguistics.

[44] OECD (2024), "Regulatory approaches to Artificial Intelligence in finance", *OECD Artificial Intelligence Papers*, No. 24, OECD Publishing, Paris, https://doi.org/10.1787/f1498c02-en.

[45] Crisanto, J.C., Leuterio, C.B., Prenio, J. and Yong, J., 2024. *Regulating AI in the financial sector: recent developments and main challenges*. FSI Insights on policy implementation No. 63. Available at: https://www.dirittobancario.it/wp-content/uploads/2025/01/Documento-di-ricerca-BIS-dicembre-2024-n.-63.pdf [Accessed 6 March 2025].

[46] OECD (2024), "Regulatory approaches to Artificial Intelligence in finance", *OECD Artificial Intelligence Papers*, No. 24, OECD Publishing, Paris, https://doi.org/10.1787/f1498c02-en.

[47] Crisanto, J.C., Leuterio, C.B., Prenio, J. and Yong, J., 2024. *Regulating AI in the financial sector: recent developments and main challenges*. FSI Insights on policy implementation No. 63. Available at: https://www.dirittobancario.it/wp-content/uploads/2025/01/Documento-di-ricerca-BIS-dicembre-2024-n.-63.pdf [Accessed 6 March 2025].

[48] Treasury Releases Report on the Uses, Opportunities, and Risks of Artificial Intelligence in Financial Services | U.S. Department of the Treasury

AIFS0030

AI expertise, data governance, model risk management, contracting with third parties, and the use of AI in regulatory compliance (regtech).

b. New Guidance in the UK Context

33.   In the 2023 Feedback Statement[49] on the Bank of England and FCA's Discussion Paper 5/22 - Artificial Intelligence and Machine Learning,[50] respondents emphasized the need for greater regulatory clarity within the UK context. They highlighted areas requiring further guidance, particularly concerning AI-driven models. Key concerns included defining bias and fairness in AI, specifically how firms should interpret the Equality Act 2010 and FCA Consumer Duty in this context. Other issues included the interaction between data protection/privacy rights and AI techniques, the explainability of AI applications, and the governance and accountability of AI-driven decisions within firms.

34.   Building on these findings, the 2024 survey by the Bank of England and FCA on AI in financial services reaffirmed these concerns.[51] Respondents cited data protection and privacy as the most significant regulatory constraint, followed by resilience and cybersecurity rules and a lack of clarity around the FCA Consumer Duty. Additionally, technical factors - such as safety, robustness, transparency, and explainability - were identified as non-regulatory barriers to AI adoption. Previous surveys of industry users have cited data corruption as a particularly pernicious threat vector of concern.[52]

35.   While these challenges persist, regulatory efforts to provide guidance on AI-related systems are not entirely new. An early example is the 2007 Financial Services Authority (FSA) statement on Automated Anti-Money Laundering Transaction Monitoring Systems (ATM),[53] later referenced by the FCA in its policy guidance. The FSA recommended that firms understand both the capabilities and limitations of their ATM systems and manage associated risks accordingly. For instance, operators were expected to explain why a system made a particular recommendation.

36.   In the context of AI, this guidance extends to the concept of explainability. As Crisanto et. al. (2024) describe, explainability refers to the transparency of an AI system's decision-making process, particularly in terms of how outputs (e.g., underwriting decisions) are derived from inputs (e.g., customer data). This includes clarifying how the system functions and makes its decisions. However, as AI systems become more complex and more capable of processing vast, diverse datasets to identify intricate patterns, their decision-making processes often become less transparent – this is known as the "black-box conundrum." As these systems achieve higher performance, they may do so at the cost of explainability, making it harder to understand how decisions are made.[54]

---

[49] FS2/23 – Artificial Intelligence and Machine Learning | Bank of England
[50] DP5/22 - Artificial Intelligence and Machine Learning | Bank of England
[51] Artificial intelligence in UK financial services - 2024 | Bank of England
[52] Kumar, Ram Shankar Siva, et al. "Adversarial machine learning-industry perspectives." 2020 IEEE security and privacy workshops (SPW). IEEE, 2020.
[53] Automated Anti-Money Laundering Transaction Monitoring System
[54] Crisanto, J.C., Leuterio, C.B., Prenio, J. and Yong, J., 2024. *Regulating AI in the financial sector: recent developments and main challenges*. FSI Insights on policy implementation No. 63. Available at: https://www.dirittobancario.it/wp-content/uploads/2025/01/Documento-di-ricerca-BIS-dicembre-2024-n.-63.pdf [Accessed 6 March 2025].

AIFS0030

37.   The FSA guidance was issued nearly 20 years ago and does not provide clear parameters for what would constitute sufficient compliance for AI models that are opaque in their reasoning. Similarly, while the FCA's guidance on firms' duties to keep consumers informed requires providers to communicate relevant information on how their data is used, the broad nature of this requirement leaves room for varying interpretations. In the context of AI, questions arise about the level of specificity needed in this communication. For instance, does this requirement include explainability? If so, what constitutes sufficient information to explain an AI system's decision-making process?

38.   **One way to address the lack of clarity and guidance in existing regulations is through the publication of transparency reports by regulatory agencies**. These reports would disclose aggregate data on how technology use is negatively impacting people, and particularly instances where automated decisions have been challenged by customers, providing valuable insights for the public. Such reports could benefit the banking and finance industries, as well as technology firms and researchers, by helping them balance innovation with consumer protection.

39.   The FCA and PRA already publish various reports and information on their websites, including annual reports.[55] The FCA also responds to individual information requests in line with the Freedom of Information Act 2000.[56] **However, additional transparency regarding enforcement actions would be beneficial**. Currently, only enforcement actions that lead to regulatory outcomes are publicised, with little insight into the agency's decision-making process during investigations. Recognizing this gap, the FCA has launched a public consultation, acknowledging that it rarely shares updates on ongoing investigations. The agency noted that by the time enforcement actions become public - such as censures or penalties - their value in providing reassurance, education, and industry guidance is often diminished.[57]

40.   It is worth noting that the FCA is studying the impact of AI in the market through regulatory sandboxes, and research engagement. **These are positive initiatives, seen as good practices for identifying regulatory gaps in the context of changing technologies, and the agency should continue to invest its resources in such endeavours** as they have the potential to help the agency shape its regulation in face of technological innovation.[58]

41.   **Therefore, whether the need for new AI specific regulations within the banking and financial sectors is arguable, it is clear that regulators should provide further guidance regarding practical implementation of existing frameworks to AI usage, including interpretation and enforcement**.

c. The Government's Regulatory Role

42.   Even though the private sector is already undertaking efforts to self-regulate, and adopting their own standards to comply with existing regulations in face of AI, public action is necessary not only for uniformization purposes (making it easier for consumers to evaluate

---

[55] Annual Report and Accounts 2023/24
[56] For example: Information on cyber security breaches reported and related investigations - June 2024 | FCA; Information on investigations into breaches under SM&CR - November 2024 | FCA
[57] CP24/2: Our Enforcement Guide and publicising enforcement investigations – a new approach
[58] Ranchordas, S., 2021. *Experimental regulations for AI: Sandboxes for morals and mores*. University of Groningen Faculty of Law Research Paper No. 7/2021. Available at: https://ssrn.com/abstract=3839744 or http://dx.doi.org/10.2139/ssrn.3839744 [Accessed 6 March 2025].

market solutions), but also to ensure that small players and new entrants are not jeopardized, as well as ensure that policy goals are duly met. A sole self-regulatory approach bears the risk of relying on industry's interests, which may not be completely aligned with the outcomes sought by the regulator. **Therefore, hybrid policy approaches are preferred, where the government plays coordinating and enforcement roles**.[59]

43.   One instructive method comes from the Monetary Authority of Singapore. In 2018, it co-created its Fairness, Ethics, Accountability, and Transparency (FEAT) principles along with the financial industry. Subsequently, it launched the Veritas Initiative, which is a consortium of government owned entities and private enterprises. The Veritas Initiative has developed assessment methodologies and assists institutions in having a verifiable method of incorporating each of the FEAT principles in their AI and data analytics driven solutions.[60] The Initiative has also launched an open-source toolkit for financial institutions to utilise. The public assessment methodologies and verifiable methods of incorporation of the principles will provide guidance to organisations in evaluating their AI based solutions and incorporating principles even if they are not a part of the programme. Further, feedback from successive rounds of testing will also assist the regulators in fine-tuning the principles and provide better guidance.

44.   Emerging research from other jurisdictions has flagged concerns regarding data quality issues in the development of systems adhering to the regulatory principles.[61] Further, even when data does exist, concerns regarding its quality: completeness, accuracy, lack of bias exist[62]. **This area may require regulators monitoring developments and creating standards in the training data quality and conditions for access to such data by emerging players or in the alternative creating incentives for financial institutions to consistently improve their training data quality while making it accessible also sharing such data with others (including terms in open systems)**.

d. Regulatory Coordination Between Agencies

45.   Considering the requirements for compliance with existing regulations may involve practices that are under consideration by other regulatory agencies, coordination between agencies is crucial to avoid overlapping regulatory burdens. In the US, the Treasury recommended that "*financial regulators continue coordinating to identify potential enhancements to existing risk management frameworks and working with other government agencies to clarify supervisory expectations on the application of frameworks and standards, where appropriate*".[63]

46.   In the UK, the Information Commissioner's Office (ICO) has already enacted more detailed orientation towards the use of AI and compliance with data privacy regulations – which continue to apply regardless of the technology employed in a cross-sector way, meaning that

---

[59] BLACK, J. (2001). Decentring regulation: Understanding the role of regulation and self-regulation in a 'post-regulatory' world. *Current Legal Problems*, 54(1), pp. 103-146.

[60] Veritas, *Trust in AI - Document 5: From Methodologies to Integration*, available at: https://www.mas.gov.sg/-/media/mas/news/media-releases/veritas-document-5---from-methodologies-to-integration.pdf [Accessed 14 March 2025].

[61] M Konrad Borowicz, The data quality problem (in the European Financial Data Space), *International Journal of Law and Information Technology*, Volume 32, Issue 1, 2024, eaae015, https://doi.org/10.1093/ijlit/eaae015

[62] Keswani, Vijay, and L. Elisa Celis. "Auditing for Diversity Using Representative Examples." Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021.

[63] Treasury Releases Report on the Uses, Opportunities, and Risks of Artificial Intelligence in Financial Services | U.S. Department of the Treasury

AIFS0030

banking and financial services providers need to abide for them in addition to sector specific regulations.

47. For example, the ICO has issued guidance related to machine learning, and explainability.[64] If the FCA and the PRA adopt different concepts and guidance regarding those topics, regulated entities might be overburdened, especially if the agencies' orientations are not aligned. In such a scenario, regulated entities will have to comply with different clashing frameworks that apply to the same use case – which is mostly harmful for new entrants and smaller market players.

48. **Ensuring consistency across agencies that oversee the banking and financial sectors is important for achieving effective compliance**. Such a concern was raised by respondents both in the US context and the UK context.[65] Accordingly, initiatives such as the Digital Regulation Cooperation Forum (DRCF)[66], and the recent joint letter of the FCA and the ICO from March 2025[67] - seeking industry input on how both agencies can collaborate with the private sector to enhance regulatory clarity and support growth - are seen as positive examples of regulatory coordination. Encouraging such efforts can help to ensure a more cohesive and predictable regulatory environment.

e. Implementing organisational measures

49. **Augmenting data governance** through multiple routes such as training, validating and testing the data that is used for model design purposes in multiple phases from the very first instance of its collection as well as later during its actual use, subject to appropriate governance practices such as data provenance[68], as well as **laying down organisational norms for dedicated documentation and record-keeping** such as maintaining a clear documentary trail for the design and development of AI models, changes made during their lifecycle, the levels of their regulatory adherence, logging and traceability of AI-incidents - these represent another set of best practices that may yield positive results in this front.[69]

f. Upscaling Regulators' Skills

50. Finally, in the same way banking and financial services providers need to leverage their board members, managers, and staff's skills to deal with the adoption of AI and the organisational changes it may bring, **public agencies too need to invest in qualifying and training its personnel to properly handle overseeing the market and enforcing its regulations.**[70] The FCA's Data, Technology and Innovation (DTI) division (created in 2021) is,

---

[64] Guidance on AI and data protection | ICO; Big Data, Artificial Intelligence, Machine Learning, and Data Protection | ICO; Explaining decisions made with AI | ICO

[65] Treasury Releases Report on the Uses, Opportunities, and Risks of Artificial Intelligence in Financial Services | U.S. Department of the Treasury; FS2/23 – Artificial Intelligence and Machine Learning | Bank of England

[66] Home | DRCF

[67] FCA and ICO letter: Supporting AI, innovation and growth in financial services

[68] Wenger, Emily, et al. "Data isotopes for data provenance in DNNs." Proceedings of Privacy Enhancing Technologies Symposium (PETS), July 2024.

[69] Chourasia, I., Ranganathan S. and Pitchandi P., 'Augmenting the reliability of AI in financial services', TCS, 2024. Available at https://www.tcs.com/what-we-do/industries/banking/white-paper/ai-financial-services-risk-governance [Accessed March 15, 2025]

[70] Crisanto, J.C., Leuterio, C.B., Prenio, J. and Yong, J., 2024. *Regulating AI in the financial sector: recent developments and main challenges*. FSI Insights on policy implementation No. 63. Available at: https://www.dirittobancario.it/wp-content/uploads/2025/01/Documento-di-ricerca-BIS-dicembre-2024-n.-63.pdf [Accessed 6 March 2025].

AIFS0030

hence, seen as an institutional improvement, while recruiting personnel with technical skill sets, including data science, AI, data governance, digital product development, cloud architecture, engineering, among others.[71]

TRAIS' researchers involved in this response:

- Isabela Parisio, LL.M, Postdoc. Research Associate at King's College London (Project Lead, UK)
- Dr. Krishna Pillutla, Assistant Professor of Data Science and AI at IIT Madras (Project Lead, India)
- Dr. Deborah Olukan, Postdoc. Research Associate at King's College London (Co-Investigator)
- Dr. Arjun Bhagoji, Assistant Professor at the Centre for Machine Intelligence and Data Science, IIT Bombay (Co-Investigator)
- Aman Gupta, Assistant Professor of Law & Technology at the West Bengal National University of Juridical Sciences (Co-Investigator)
- Dr. Shouvik Kumar Guha, Associate Professor of Law & Technology, at the West Bengal National University of Juridical Sciences (Co-Investigator)

About Responsible AI UK

Responsible AI UK brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society. It is an open, multidisciplinary network, drawing on a wide range of academic disciplines. This stems from our conviction that developing responsible AI will require as much focus on the human, and human societies, as it does on AI. Funded by the Technology Missions Fund, we convene researchers, industry professionals, policy makers, and civil society organisations.

About CeRAI India

CeRAI is a virtual interdisciplinary research centre at IIT Madras. The centre envisions to be a premier research centre for both fundamental and applied research in Responsible AI with immediate impact in deploying AI systems in the Indian ecosystem. The centre aims to pursue research in the domains of Ethical and Responsible AI and become the standard body in the country to recommend guidelines and policies to make deployable AI models/ systems more accountable, explainable and responsible.

April 2025

---

[71] FCA Response Letter